

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:20:59 ; Search time 10173.5 Seconds
(without alignments)

13311.631 Million cell updates/sec

```

Title: US-09-171-553B-3
Perfect score: 8209
Sequence: 1 gtgtgtcagactgtggcc.....aaaaaaaaaaaaaaaaaa 8209

```

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size :

Total number of hits satisfying chosen parameters: 2944280

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Listing first 45 summaries

Database :

GeneEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8209	100.0	8209	6	A66553	A66553 Sequence 3
2	7751	94.4	8196	6	A66552	A66552 Sequence 2
3	4833	58.9	8918	14	PEN133816	AJ133816 Porcine e
4	4640	56.5	8849	14	PEN133817	AJ133817 Porcine e
5	4540	55.3	8763	14	PEN293657	AJ293657 Porcine e
6	4532	55.2	8750	14	PEN133818	AJ133818 Porcine e
7	4096	49.9	7808	14	PER17013	Y17013 Porcine end
8	2967	36.1	7333	6	AF038601	Sus scrofa
9	2967	36.1	7333	6	AR130474	AR130474 Sequence
10	2682	32.7	3320	6	A66551	A66551 Sequence 1
11	2034	24.8	3025	6	AF147808	AF147808 Sus scrof
12	1971	24.0	1971	14	AF014162	AF014162 Porcine e
13	1499	18.3	1974	6	A66559	A66559 Sequence 9
14	1466	17.9	3482	6	AX002804	AX002804 Sequence
15	1466	17.9	3482	14	PEREN2	Y12229 Porcine end
16	1064	13.0	8918	14	PEN293656	AJ293656 Porcine e
17	876	10.3	927	4	SSU7599	U77599 Sus scrofa
18	848	10.3	4402	6	AX052635	AX052635 Sequence
19	797	9.7	6076	6	AX052636	AX052636 Sequence
20	797	9.7	7352	6	AX052634	AX052634 Sequence
21	746	9.1	6076	6	AX052647	AX052647 Sequence
22	704	8.6	8132	6	AR130475	AR130475 Sequence
23	622	7.6	942	14	MLVRT10	X93933 Murine Leuk
24	570	6.9	4918	6	AX052637	AX052637 Sequence
25	557	6.8	8132	6	AF038600	AF038600 Sus scrof
26	547	6.7	7873	6	AX052638	AX052638 Sequence
27	457	5.6	2452	6	AX002802	AX002802 Sequence
28	457	5.6	2452	14	PEREN1	Y12228 Porcine end
29	446	5.4	1042	14	PERY17012	Y17012 Porcine end
30	437	5.3	2000	14	PEN288588	AJ288588 Porcine e
31	375	4.6	8060	6	AR130473	AR130473 Sequence
32	375	4.6	8132	6	AF038559	AF038559 Sus scrof
33	367	4.5	1968	14	PEN288589	AJ288589 Porcine e
34	335	4.1	2000	14	PEN288592	AJ288592 Porcine e
35	321	3.9	746	14	PEN298073	AJ298073 Porcine e
36	318	3.9	590	14	PEN298074	AJ298074 Porcine e
37	276	3.4	276	4	SSA5400	AJ005400 Sus scrof
38	225	2.7	276	4	SSY18745	Y18745 Sus scrofa
39	223	2.7	276	4	SSA55409	AJ005409 Sus scrof
40	217	2.6	276	4	SSY18748	Y18748 Sus scrofa
41	201	2.4	276	4	SSY18749	Y18749 Sus scrofa
42	189	2.3	276	4	SSA55410	AJ005410 Sus scrof
43	161	2.0	810	14	AF033359	AF033359 Porcine e
44	158	1.9	522	14	AF033360	AF033360 Porcine e
45	158	1.9	927	4	SSU77600	U77600 Sus scrofa

[illegible]

CDS

588..2162
/codon_start=1
/product="VIRAL CORE PROTEIN"
/protein_id="CAB39347.1"
/db_xref="GI:4538107"
/translation="MGQVTPTLSLTLIDHMTVEVRSRAHNLVQYKKGPMPOTFCSEMP
TFDVGWPESEGFENSEIILAVKAILFQTPGSHPOEPEYILTMODLADPPPMKPMIN
KPRKGRPILALGKRNKSAKEVPESSSYLPRDGAADLADGTPGSPPLSSGCCG
TSAPGPAVVEGPAAGTGRSRGATPERDEALILPLRYGPPMAGGLOLOLTERERILIEA
ADLYNMKTNIHPPSESDPOLRLGLVESLMPFSHPMTDCOOLLOLTLTERERILIEA
RKVPGADGRPTOONELDMQFPLTRGWMYNTAEGSESLTKYKQALVAGLKGASRRP
TNLAKVEVMQGNPPSPVLELMEAPFRFTPDPESEOKASVALAFGOSALDR
KKILOGLQDAELRLDLYREAEKYYRRETEEEKQEKREEREREEREROREKNT
KILAAVEGSSRRERDRFRKIRSGPDSGNLGNRPFLDKQCAVCKEKKHMANRCPK
KGNKGPVLALEKCD"
<2163..5747
/codon_start=1
/product="REVERSE TRANSCRIPTASE"
/protein_id="CAB39348.1"
/db_xref="GI:4538108"
/translation="GRGSDPLPEPRVTLKVEGQPEVLDVCAESVYLQPLKAKE
KSMVMATGROYPMTTRRTVDLGYCRVHSLVLPCEPVLPLGLDKLTKMAQIST
EOGRPEVSNKRPITVTLQDDEYRLISQVAPDDIDISLQEPQAAEAGMLA
KOYPOVIOLKASATPVSVKQYPLSEARGILVPHVQRLIOGILPVOSPMWTLPLP
VRKPTNDYKPVODLRVNRKVDIHPVNPYNLISALPERMWTVIDLDAPECL
RLHPTSQPLFAFEMDPDGTGRTGQLTWRLPOGFKSPITLFDALHRLDANRHOPO
VTLLQYVDLLLAGATKODLEGTALLESLSIGRASAKAOCIRREVTYLGSLR
GGORMLTEARKTVOIIPATTAQVREFLGTAGRCRMLPGRATLAALYLTLEKQ
GFSMAPEHOKAPDAIKKALISAPALAPVYTRFTLYVERKGVAGVLTQILGEMR
PVAYLSKIKDPAVSGMPCVCLKALIAVALIKADKLTLLQNTVLAVALENIVQRP
DRMTNARKHTVOSLLTERVFAPPAALIPATLLPEETDEVTDHQOLIEETGVR
KDLTDLPLGEVLTWPTDSSVVEGKRNAGAAYVGGRTIWAASSPESTSOAKEM
ALTOALRLAEGESINITYDSRYAFATAHVAGIAYKORGLTSGAREIKREILSLE
ALHLPKRALIHCPOHOKADLISRGMOADRYAQAQAVULLPILEPKAPEERQ
YTLDEMOELIKIDIOSEPEGTCTCYGKELPHKEGLEYVQIHLTHLGLKHLQOL
VTPSPHYVRLRGVADSVKRCPCOLVANSPRIPEGRILSHGHWEDLETK
PAKYGNKYTLVEVDPSGWEAVYPTKKESTVYAKILIEITLPRGIGSDMGPA
FAVQVSQGLAKILIGDKLHCAYRPOSSGOVERNMTLIEITLKTTEGSDIMWALL
PEVLEFRVNTPCQFGLPYELLYGPPPLAIFAFMSAPLADLSLOPLRKALEVRQ
RAMKOLREAVSGSDLOVPHRFQVGDVSVYRHRACNLETRMKPGLVLLITPTAVKE
GIPTWIIHASHVAPAPPPDSGMARKEKTEMLKRLRLVPIYSNNNSPGQ
5620..7590
/codon_start=1
/product="VIRAL ENVELOPE PROTEIN"
/protein_id="CAB39349.1"
/db_xref="GI:4538109"
/translation="MHTLSRRLPTRGCEPRRLRIPSPASIAFLTLITTPQASK
RLIDSNRHPRLSTLWLIIDPTGTVTVSTGVARGTWMLHFLCLMLINPAKSTP
PULVSYGYCCPTEKERTCGGSESCSRMSCVTSNDGMKWPISLDIVKESFVN
SGPKYKMKMLKDKSCSPSDLYLKLISFTEKRTGKSKVKNWELISDLYGGGAG
STLTLRLIETGTEPPVAMPDPKVLAEQGPALPPHNLVPOLSLRDPITOPSPNS
ITGLIPTNTPRNSPGVPVKTGORLFSLIQAFQAINSDPAPATSCMCTSSGPPYE
GMAKERKFNVTKEHNOCTGMSRNKLATLEVGSGKCVKQAPRHQHCYSTVYEOA
SENOYLPGYNRMACNTGLPCVYSTVENSQKPCVAVQIIPVRYTHPEVILDEVD
YRNRNPREPYSITLAVMLGTAAGVGTGTALITGQLEKGLGELHMAEMTDLA
LKESYNSLEBSLTSLSEVLDNRGLDLFLREGLLAALKECCFVVDISGALRDSM
NKLKRLKERRRRERADQGMENGFNNSPMWTLLSALTGLVLVLLLTITVGPLLINR
FVAFVREVSVAVQIIMVLROQYQGLISQENDL"

CDS

BASE COUNT 2168 a 2064 c 2154 g 1823 t
ORIGIN
Query Match 100.0%; Score 8209; DB 6; Length 8209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtaggtacgaactgtgagcccaagcgactggaataaaatcctctgtgtgattgcaac 60
DB 1 GTGGCTGACGACTGTGGGCCCCAGCGCTTGGAATAAAATCCTCTGTGTTGCATC 60
QY 61 aagaccgtctctgtgaatgatttgagggtgtcgcccttcagagcccgagagggagat 120
DB 61 AAGACCCTCTCTCGATGATTGGGGGTGTGCGCTTTCGAGACCCGAGAGGGGGGAT 120

CDS

QY 121 tgtctttacttgaccttcatctgtgtcgttgccggaaatccctgcagacacctta 180
DB 121 TGTCTTTACTTGACCTTTTCATTTGGTGTGGCTGGCCGGGAAATCCTGCACACCTTA 180
QY 181 caccggagacagcacttgagagtgaaaggaatccctcttgaaagctgtgtgtcgcc 240
DB 181 CACCGGAGAACCGACTTGGAGGTAAAGGGATCCCTTTGGAACGTGTGTGTGTGGCC 240
QY 241 ggcgtctctgtctgaatgtctgttttgagtgatgcccgtttgcagctgtcct 300
DB 241 GGCCTCTCTGTCTGAATGCTCTGTTCGTTGATGCGCGCTTTCGAGCTGTCTCT 300
QY 301 ctaagacgtaagagacttgagagctgtgatacaagacgtgtctaagagatcaagagct 360
DB 301 CTCAGACCGTAAGAGACTGGAGAGCTGTGATCAGACACAGCTGCTAGAGAGATCAGAGCTG 360
QY 361 ccaaccttgaggagagcccgaggaggtgaggagagccagagcactgtgtctact 420
DB 361 CCAACCTTGAGGAGAGCCCGGAGGTGGGAGAGCCAGGAGCGCTGTGTCTCTACT 420
QY 421 gtcgacagagagcagactctgtgttgaaagaaagctccctccgagccgtcg 480
DB 421 GTGCGTACAGAGACCGAGTCTGTGTTTGAACGGAAGGCTTCCCTCGCGCGCTCG 480
QY 481 actctttgacctgtgttgaaagacgagcggtcgctgtgtctgatacgtttgttt 540
DB 481 ACTCTTTGCTGCTGTGTGGAAGACGCGGAGCGGTCGCTGTGTGATCTGTTGTGTTT 540
QY 541 cgtctcgtgtctcttctgtctgtctgtctgtctgtctgtctgtctgtctgtct 600
DB 541 CGTCTCGTGTCTCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 600
QY 601 tgactaccccccttgcttgactctcgacatcttgacgtgaagttagatccagggtcata 660
DB 601 TGACTACCCCCCTTGCTTGACTCTCGACATTTGACTGAAAGTTAGATCGAGGCTCATA 660
QY 661 atttgcagttcaagtttaagaagggaccttgcagacttctgtccctctaaatggcaca 720
DB 661 ATTTGTCAGTTCAAGTTTAAGAAAGGACCTTGCGAGACTTCTGTGCTGTCAATGGCCAA 720
QY 721 catcagatgttgatagggcaccacagagggagaccttaattcgaataatccctgtgta 780
DB 721 CATTCAGATGTGATAGGGCACCACAGAGGGAGACCTTAAATTCGAATTAATCCTGCTGTTA 780
QY 781 aggaatcatatttcaagacttgagcagcgctctcaatccctgacagagccctatactcta 840
DB 781 AGGCAATCATTTTCAAGACTTGAGCCGCGCTCTCATCTGATCAGAGACCTTATATCTTCA 840
QY 841 cgtgcaagatttggcagaagaatccctccgcacatggtttaaacatctgtaataaacaa 900
DB 841 CGTGCAAGATTTTGGCAGAAAGATCCTCCGCAATGCTGTTAAACCATGTGCTAAATAAACAA 900
QY 901 gaaagcaaggtccccgaatccctgtctcttgaggagaaacaaactctgcccgaataag 960
DB 901 GAAAGCAAGGTCCCCGAATCTCTGCTTTTGAAGGAAACAAACAATCTGCGCGGAATAAG 960
QY 961 tgaagccctctctctgtatctaacccagagatcgaagagcgccagactgtgcccgaaccc 1020
DB 961 TCGAGCCCTCTCTCTGTATCTAACCCGAGATGAGAGAGCGCCGACTTGCCGGAACCC 1020
QY 1021 caactgttcccccaaccccttatcagacagaggtgtctgttgaggagactctgtccctc 1080
DB 1021 CAACCTGTCTCCCAACCCCTTATCAGACACAGAGGTGTCTGTGAGGGAGACTGTGCCCTTC 1080
QY 1081 ctgagactcgtgtgttgaggagacactgtctgcccggagactcggagagcgccacccc 1140
DB 1081 CTGAGACTCGTGTGTGTGAGGAGACCTGTCTGCGGAGCTCGGAGAGCGCCACACC 1140
QY 1141 cgaagcgagacagagatcgatatactacgctgtgcacactatggccctccatgagcag 1200
DB 1141 CGAGCGGAGACAGAGATCGGATATTAACGCTGTGCGACACTATGGCCTCCCATGCGAG 1200

Db	3361	AGGCACGGAGAAGAAAACGTGATGTCACAGATACCGGGCCCCAACACACAGCCAAACAAAGTGAAGAC	3420
QY	3421	agtttttggggagcagctcggatatttgcagactctgtgatacccgagggtttgcgacctagcag	3480
Db	3421	AGTTTTTTGGGACACAGCTGGATTTTGGCAGACTGTGGATGCCGGGGTTTTGGACCTTAGCAG	3480
QY	3481	ccccacttaccgctaacccaagaagaaggaggtattccctcgggctcctctgagccaga	3540
Db	3481	CCCCACTCTACCCGCTTACCCAAAGAAAAGGGGGATTCTCTGGGCTCTTGACACCCAGA	3540
QY	3541	aggcaattgatgtatcaaaaagggccctctgagcgacctgctcttggccctccctgacg	3600
Db	3541	AGGGATTGTGAGTCTATCAAAAAGGCCCTGCTGAGCCAGCACTCTCTGGGCCCTCCCTGACG	3600
QY	3601	taacaaacccctttacccttattgtatgtatgagtgaaggaatagccggaggattttaa	3660
Db	3601	TAACTAAACCCCTTTACCTTTATGTGTGATGATGAGTGTAAGGGAATTAACCCGAGAGTTTAA	3660
QY	3661	cccaaacaccttagagacaatggagagagacctgtccctaccctgtcaagaagactgtacctg	3720
Db	3661	CCCAAAACCCCTTAGAGACATGGAGAGGAGACGTGTTGCCACTGTCAAGAAAGCTTGATCTCG	3720
QY	3721	tagccaatgtgtgttggcccgatctgtctgaagctatcgcagctgttggcatalactgttcaag	3780
Db	3721	TAGCCAGTGTGTGGCCCGGTATGTGCTTAAGGCTATGCACAGCTGTGCCATCTGTGCATAGG	3780
QY	3781	acgctgacaatltbaacttttggacaagataatacgtatataagccccatcatttggaga	3840
Db	3781	ACGGTGACAATTTGACTTTTGGGACAAATTAATTAAGTAAAGTAAAGCCCCCATGATTTGACGA	3840
QY	3841	acatcgtcttggcagccccacagaccgaatggatgaaccaagcccgcatgaaccacatacaaa	3900
Db	3841	ACATCGTTTCGGCACCCCCAGACCGATGGATGATGATCAACAGCCCGCATGACCCACATATCAAA	3900
QY	3901	gcctgccttcaacgaagaaggttaacttctgcaccacagccgcttcaacccttcaactc	3960
Db	3901	GCGTGCCTTTCACAGAGAGGGGTACTTTTGGCTCCACACAGCCGCTCTCAACCTTCCACCTC	3960
QY	3961	ttctgccttgaagagactgtatgaaccagtgactaatgatctgtccatcaaatattgtatbag	4020
Db	3961	TTTCGCTTGAGAGAGACTGATGAACCCAGTACTCAATGATGGCATCAATTAATTTGATGAGG	4020
QY	4021	agacttgggtgcgcgaagagaccttaacagacataacgctbaacttgcgggaagtgttaaccgt	4080
Db	4021	AGACTGGGGTCCCGAAGACACTTACAGAGATACCCCTCACTGGAGAAATGCTTAACCTGGT	4080
QY	4081	tcactgaacggaagacgtatgttgttgaaggtgaagaagatagtcttggggcgcgatgttgg	4140
Db	4081	TCACTGACGGAACAGACTATGTGTGTGAAGGTAAGAAGATGGCTGGGGGCAATGTGTGG	4140
QY	4141	acgggagaccgcgaagatctgtggccagagaccttgcggaaggaacttcagcgcaaaagctg	4200
Db	4141	ACGGGACCCGCAAGATCTGTGGCCAGACGCTTCCGGAAGAAATTCAGCGCAAAAGGCTG	4200
QY	4201	agctcatlgycccttcaacgacaagcttltgcgcttgcgcgaagggaaaatccataaacttata	4260
Db	4201	AGCTCATGGCCCTCACGACAAGCTTTGGCGCTGGCGGAGGGAAATTCATAACATTTATA	4260
QY	4261	cggagcagcagatgatgccttltgcgactgcacgctacacgtygggcacatataaacaagg	4320
Db	4261	CGGACACACAGTATGCTTTGGAGCTGCAACAGTACAGGGGGCATCTATTAACAAAGGG	4320
QY	4321	ggttgccttacctagcaggaagggaataataaagaagaaggaaattctaagcttlttg	4380
Db	4321	GGTTGCTTACCTACGAGGAGGGAATTAAGAACAAAGAGGAATTTCAAGCCTATTAG	4380
QY	4381	aagccttaacttgcacaaaaggctagatbatatacacttgccttgcacatagaagaagca	4440
Db	4381	AAGCCTTACATTTGCCAAAAGAGCTAGCTATTAATTAACATGTGCTGGACATCAGAAAGCCA	4440
QY	4441	aagatcctcatcttaaggaaacagatagctctgacccgggttggccaagcagagcccgag	4500
Db	4441	AAGATCTCATATTTAAGGGAACCAAGATGGCTGACCGGGTTGGCCAAACAGGACGCCAG	4500

QY	4501	ctgttaaaccttcgtccctaaatagaacacgcccacaaagcccaagaccacaaagaacgtaca	4500
Db	4501	ctgtttaccttcttgcccttattatataaaacgcccaaaagcccaagaccacaaagaacgtaca	4500
QY	4561	cccttaagaagacttgcgaagaagaataaaaaagaataagacagttctctcgtagactccgagagga	4620
Db	4561	cccttaagaagacttgcgaagaagaataaaaaagaatagaccagttctctcgtagactccgagagga	4620
QY	4621	ccctgtcataacctatgtatggaagaatactctgcgcccaaaaagaaggttagaatactcc	4680
Db	4621	ccctgtcataacctatgtatggaagaatactctgcgcccaaaaagaaggttagaatactcc	4680
QY	4681	aacagatacatcgtctaaacccacccaagaaactaaacactcgaagcagtttgcgaacat	4740
Db	4681	aacagatgacatcgtctaaacccacccaagaaactaaacactcgaagcagtttgcgaacat	4740
QY	4741	cccctatacatgtcttcgaagctaacagaagttgctcgtactcgglygtlcaaacattgtlgc	4800
Db	4741	cccctattacatggttctcgaagctaacagaagttgctcgtactcgglygtlcaaacattgtgnc	4800
QY	4801	ccctgcacgcttggttaatgcttaacaccttcacgaataacctcccaagaagaagacttaagggaa	4860
Db	4801	cctgcacgcttggttaatgcttaacaccttcacgaataacctcccaagaagaagacttaagggaa	4860
QY	4861	gccaccccaagcgcctcacttgcgaagttgcacttcaatbaggtaaaagccgactaaataagaa	4920
Db	4861	gccaccgccagcgcctcacttgcgaagttgcacttcaatbaggtaaaagccgactaaataagaa	4920
QY	4921	acaaatactctatgttttttgttagacaaccttttcagagatgggttagaggctatactccta	4980
Db	4921	acaaatattctatgttttttgttagacaaccttttcagagaggggttagaggctattatccta	4980
QY	4981	agaagaagacttcaacccgttgctgcgaagaataactcgtgagaagaattttccaagacttg	5040
Db	4981	agaaagagacttcaacccgttgctgcgaagaataactcgtgagaagaattttttccaagatttg	5040
QY	5041	gaatacctaaggtataatgggttcagacaatgtgtccagcttcgtgtgccaggtlaagtcag	5100
Db	5041	gaattaccttaaggtatattgggttcagacaatgtgtccagcttcgtgtgccaggtatagtcag	5100
QY	5101	gacttgcacaagatactggagatctgcactctggaactcagttgcacataagaccaccaagct	5160
Db	5101	gacttgcgcccaaatattttggggattgatttggaamaactgcatgtgtgcatacagaccctaaagct	5160
QY	5161	caggacaggttagagagaatgaatagaaccatlaaagaagacccttaaccaattgaccaag	5220
Db	5161	caggacaggttagagagaagatgaatagaaccatlaaagaagacccttaaccaaatgtgaccag	5220
QY	5221	agactggcattaaatgatcttgaatgtgctctcctgcaccttgccttttaaggltgaaggaca	5280
Db	5221	agactggcattaaatgatcttgaatgtgctctcctgcaccttgccttttaagggtgagagaca	5280
QY	5281	ccccttgcacagtttgggcttgaacccccctacgaaattgtcctacagggggaaccccccgcttg	5340
Db	5281	ccccttgcacagtttgggcttgaacccccctattgaattgtcctacaggggggaaccccccggttg	5340
QY	5341	cagaaattgcctttgtcaacatagtgtcgtatgtgcctgccttccagccttgttctctaggc	5400
Db	5341	cagaaattgtcctttgtcacatagatagtgctgtatgtgcctttccacaccttggcttcttaggc	5400
QY	5401	tcaagagcctcagatgtgggttgaagcaacgaggttgaagaagcctcgcgggaagcctacacag	5460
Db	5401	tcaagagcctcagatgtgggttgaagcaacgagccttgaagaagcctcgcgggaagccttactcac	5460
QY	5461	gaggagaacttgcagaattccacalcgctctcacaagtgtgagattcagatcattgtatagaagc	5520
Db	5461	gaggagaccttgcagaattccacalcgctctcacaagtgtgagattcagttatgttatagagcc	5520
QY	5521	acgcttgaagaagaacctcgaagactcggltgaagaggaaccttactcgttactcttgaacaac	5580
Db	5521	acgcttgaagaagaacctcgaagactcggltgaagagggaccttattctcgtattctttgaacacac	5580

QY 5581 caagcgtgtgaaagtcgaagaatcccccctgcatcctacccacgttaagccg 5640
|||||
Db 5581 CAAGGCTGTGAAGTGAAGGAATCCCACTGGATCCATGATCCCAAGTTAAACCGG 5640
QY 5641 cgcacactcccgactcgggttgagagccgaaaaagactggaatccccccttaagctcgc 5700
|||||
Db 5641 CGCACACTCCCGAATCGGGGTGGAGACCCAAAAGACTGAGATCCCTTAAGCTTCGCC 5700
QY 5701 tcaatcgctcgttcttacttacttaacaatacctcccaagccagtgatgaagccttata 5760
|||||
Db 5701 TCAATCCCTGCTGTTCTTACTCTTAACATTAATCCCAAGCCAGTAAACGCTTATA 5760
QY 5761 gaacgctcgaaccccaatagaccttaaccttaacctggtgattttgaacctgtag 5820
|||||
Db 5761 GACAGCTCGAACCCTTAACCTTTATCCCTTAACCTGGGTGATTTAGACCTGATAGC 5820
QY 5821 ggtgcatctgtaaaagcactcgaggtgtgtctcctgaagagcactgtgtgacctgaact 5880
|||||
Db 5821 GGTGTCACTGTAATATACACTCGAGGTGTGCTCTAGAGGACACTGTGTGCTGAACCTG 5880
QY 5881 catttcgctccgattgattaaaccccgctgttaaaagcactcccaacctgaagtcgt 5940
|||||
Db 5881 CATTTCTGCTCCGATTGATTAAACCCGCTTTAAAGCAACCTCCAACTAGTCCGT 5940
QY 5941 agttaaggtctctatgtctcccgagcagagagaagaatactgttgagggtctcgg 6000
|||||
Db 5941 AGTTATGGGTTCTATGCTGCTCCGAGCAAGAGAAAGAAATATCTGTGGGGGTCTCGGG 6000
QY 6001 gaatcctctgtagagaatgagactgtgtcaactccaaagatggaactggaatgagccg 6060
|||||
Db 6001 GAATCTCTCTGTAGAGATGAGAGCTGCTGCTCACTCCAAAGATGAGACTGGAATGCGCG 6060
QY 6061 atctctctccagagccgggttaaaatltctcttctgtcaatcccgccggagcaagtaaaa 6120
|||||
Db 6061 ATCTCTCTCCAGGACCGGGTAAATCTCTTGTCAATTCGCGCCGGCCAGAGTAAATA 6120
QY 6121 atgtagaactataaagaatagagctgtccccaacagactagatacttaaaagata 6180
|||||
Db 6121 ATGATGAACATATATTAAGATAGAGCTGCTCCCATCAGACTTAGTTATCTTAATAATA 6180
QY 6181 agttcactgaaagaaacagagaaatlatcaaaagtgaataatgtgtatgagctgagg 6240
|||||
Db 6181 AGTTCACTGAAGAGAAACAGGAAATATTCAAAGTGTGATTAATGATGAGCTGGCG 6240
QY 6241 aatagtttttatatatgtcgaggggagccaggttccaccttaacctgtgaccttagata 6300
|||||
Db 6241 AATAGTTTTTATATATAGGGGGGAGGACAGGCTCACTTAACCATTCGCTTAGSATA 6300
QY 6301 gagaagggagacagacccctgtgtgcaatgagaccgataaagtaactgtgagcaaggg 6360
|||||
Db 6301 GAGAGGGGAGACAGAACCCCTGCTGGCAATGGGACCCGATTAAGTAACTGGCTGAACAGGG 6360
QY 6361 ccccgagccctgagacacagcataactgtccggtgtccccaatlaacctgtgagcc 6420
|||||
Db 6361 CCCCCGGGCTGAGAGCCACCGCATTAATGCTGGTCCCAATTAACCTGCTGCTGGCCCT 6420
QY 6421 gacataaagcccgccgactgaacagagtaacactgtattgttcttactaaagccataga 6480
|||||
Db 6421 GACATTAACACAGCCGCTAGACAGATGACACTGATTTGATCTTACCAACAGCCCTAGA 6480
QY 6481 aactcccaaggtgtctcgtttaagaagagacagactctcaatcctcatcccaagggact 6540
|||||
Db 6481 AACTCCCAAGGTTCTCGTTAAGACAGGACAGAGACTCTCACTGCTCATCCAGAGGACT 6540
QY 6541 ttccaagcactcaactccacagaccctgaatgcacacttcttctgttgcttctctatcc 6600
|||||
Db 6541 TTCCAAGCCATCAACTCCACAGCCGCTGATGCCACTTCTTCTGTGGCTTTGCTATCC 6600
QY 6601 tcaaggtcctcttattatgaagagatggtctaaagaaagaaatcaatgtatccaaagag 6660
|||||
Db 6601 TCAGGGCTCTCTTATTATGAAGGGATGAGCTTAAGAAAGAAATTCATGTGATCAACAAAG 6660
QY 6661 catagaataatcaatgtacatgggggtcccgaaataagcttaccctcaactgaagttccgg 6720
|||||

Db 6661 CATAGAAATCAATGATGAGGGGTCCGAAATAGCTTACCTCACTGAAAGTTCCGGG 6720
QY 6721 aaggggacatgtagatgggaaagcttcccccattcccaacacacttgcctatagctgtg 6780
|||||
Db 6721 AAGGGGACATGATAGGAAAGCTCCCATCCACACACACTTGTCTATATGATCTGTG 6780
QY 6781 gttatgaagcagccctcagaataatcagatattagactgtgttataaaggtgtgagca 6840
|||||
Db 6781 GTTTATGACACAGGCTCAGAAATCAGTATTATGATGCTGTTATTAACAGGTTGGGCA 6840
QY 6841 tgcnaacttgggttaacccccgtgtgttccacctgaagttccttaacccaatccaaatttc 6900
|||||
Db 6841 TGCNAATGCTGGGTAAACCCCTGTGTTCACCTCACTGATCTTCAACCAATTCAAAGATTTC 6900
QY 6901 tgtgtcatgttccaaatgttcccccaggtgttaccatccatctgaggaagttgtctgat 6960
|||||
Db 6901 TGTGTATGTGTCAAATGCTGCTCCCGAGTGTACTACATCTGAGAGAACTGCTTGTAT 6960
QY 6961 gaataactatcgtatlaacccgacccaagaagaccgtatcccttaacctagctgta 7020
|||||
Db 6961 GAAATGACTATGCGTATACCGACCAAAAGAGAACCCGTATCCCTTACCTAGCTGTA 7020
QY 7021 atgtcgtgactaggaagccggttggcgttagaagcaggaagcagctgtgacctgaacagga 7080
|||||
Db 7021 ATGCTCGATTAAGGAGCGGCGCTGTGAGAACAGGAGACAGCTGCTGATCAACAGGA 7080
QY 7081 ccaagcactatgagaagaagactgtgtgactacatgagccatgacagaagatctccga 7140
|||||
Db 7081 CCACACACTGAGAGAAAGACTGTGTGACTACATCCGCGCATGACAGAAATCTCCGA 7140
QY 7141 gctttaaagagctgtttagcaacttagaagagctccctgaactcttgttctgaagtgtt 7200
|||||
Db 7141 GCCTTAAAGAGCTGTGTAACACTGAAAGAGTCCCTGACTTCTTGTCTGAAGGTT 7200
QY 7201 ctacagaacccgagagagatagatctgtcgttcttaaggaaggtgtgtatgtgcagcc 7260
|||||
Db 7201 CTACAGAACCCGGAGGAGTATGATGCTGCTTCAAGAGAGGAGTGTATGTGACAGCC 7260
QY 7261 ttaagaagaatggtgtcttcttcttcttcttcttcttcttcttcttcttcttcttctt 7320
|||||
Db 7261 TTAAGAAGAATGTTGCTTCTTATGATGATCACTCGAGGACATCGAATCCATGAAAC 7320
QY 7321 aagcttagaagaaagtttagagagcgtcgaagggaaagagagctgtgacagaggtgtt 7380
|||||
Db 7321 AAGCTTAGAAGAAAGTTAGAGAGCGTCGAAGGAAAGAGAGCTGACAGGGGTGTGT 7380
QY 7381 gaagatgtgttcaacaggtctcctgtgtagaacacccctgtcttcttcttcttcttctt 7440
|||||
Db 7381 GAAGGATGCTTCAACAGGCTCCTTGATGACACCCCTCTTCTCTGTGACGGGGCC 7440
QY 7441 cttagtgcctgtctcgttacttaacgttggcccttcttcttcttcttcttcttcttctt 7500
|||||
Db 7441 CTAGTATGCTGCTCTCTCTGTTACTTACAGTTGGGCTTCTCTTATTAATAGTTGTGTGC 7500
QY 7501 ttgtttagaagacagttgagtgagtcagtcagatcaatgagtaagcagacagtaacaaagc 7560
|||||
Db 7501 TTTGTTAGAGAACGAGTGAATGATGACATGATGATGATGATGATGATGATGATGATG 7560
QY 7561 ctcttgagccaaggaagaactgaactctgaccttccagttcttaaataggactattaa 7620
|||||
Db 7561 CTCTTGAGCCCAAGAGCAAAATGACCTCTTACCTCTTCCAGTTCTTAATTAAGAACTTTA 7620
QY 7621 caagacaagaagtgagggaatggaagagtgaaatgtcaacctaaacctcccaagaccag 7680
|||||
Db 7621 CAAGACAAGAGTGGGGAATGAAGAGATGAAGTGAACCTTAACCTCCCAAGAACCCAG 7680
QY 7681 aagttataaagacttlaaagccccgaatlaacagaccctgtgtgtcgaagtaata 7740
|||||
Db 7681 AAGTTATTAAGAGCTTAATATCCCCGAAATTAACAGACCTGCTGCTGCTGCTGCTGCT 7740
QY 7741 ggtagaaggtcacacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 7800
|||||

Db 7741 GGNAGAGGTCACTTCTTATGTTCCAGGGCCGTATCTGCGCTTAAGTAACTAAC 7800

Qy 7801 aggaatagtgactaactcgtcttctggaattcgttaaaactgactgacacatagaag 7860

Db 7801 AGGAATAGTGTGACTAATCCGCTTATCTGATTTCTGTAACCTGACTGACACCATAGAG 7860

Qy 7861 aattgataacatagacagccctagtagaacatctacacacgcaatcgtcaactcgtcc 7920

Db 7861 AATTGATTAACATTTGACAGCCCTAGTACATCTCAATGCAATCTGTCACTCTGCC 7920

Qy 7921 aggaagccacagatagtgagacccctgagactatcttaaatgattgttccagagagc 7980

Db 7921 AGGAGCCACAGATAGGAGACCTCCGAGCATTTTAAATGATGTGTCACGAGAGCC 7980

Qy 7981 gggctctgataatttaaaatgattgtgtccabgagagcgagctctgataatttaaat 8040

Db 7981 GGGCTCTGATATTTTAAATGATTGTGTCAATGAGCGCGGCTCTCATTTTAAAT 8040

Qy 8041 gattgatttgaacagacagacttgttgaaccccaataaagctgtccgattccgca 8100

Db 8041 GATTGATTGTGACGACAGGCTTTGTTGTGAACCCCAATAAAGCTGTCCGATTCCGCA 8100

Qy 8101 ctggagcgagactcctctacccctgctggtglaagactgtgggcccagcgagctgtg 8160

Db 8101 CTGGGCGGCGAGTCTCTACCCCTGGGTGTACGACTGTGGGCCCGACGCGCTTGG 8160

Qy 8161 aataaaatcctctgtctgttgcatacaaaaaaataaaaaa 8209

Db 8161 AATAAAATCCTCTTGTCTGTGTGATCAAAAAAATAAAAAA 8209

RESULT 2

LOCUS A66552 8196 bp DNA PAT 29-MAR-1999

DEFINITION Sequence 2 from Patent WO9704167.

ACCESSION A66552

VERSION A66552.1 GI:4538105

KEYWORDS

SOURCE porcine endogenous retrovirus.

ORGANISM porcine endogenous retrovirus.

REFERENCE 1 (bases 1 to 8196)

AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.

TITLE PORCINE RETROVIRUS

JOURNAL Patent: WO 9740167-A 2 30-OCT-1997;

FEATURES

source

1. .8196

Location/Qualifiers

/organism="porcine endogenous retrovirus"

/db_xref="taxon:61673"

BASE COUNT 2165 a 2061 c 2146 g 1820 t 4 others

ORIGIN

Query Match 94.4%; Score 7751; DB 6; Length 8196;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 8191; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 13 tctgggccccagcgcttggaataaaatcctctctgtctgttgcatacaagcgtctct 72

Db 1 TGTGGGCCCCAGCGCTTGGAATAAAATCCTTGTGCTTGTGATCAAGACCGCTTCT 60

Qy 73 cgtgagtgattgggtgtctgctctctccagcccgagcgagggattgttcttact 132

Db 61 CGTGAAGTATTTGGGGTGTGCTCTTCCGAKCCGAGAGGGGAGTGTCTTTACT 120

Qy 133 ggccttcatttgctgcttgccggaataactctgtgacaccccttaacccagagacc 192

Db 121 GGCCTTTCATTGTGTGCTGTGGCCGGAAATCTGTGACACCCCTTACACCCGAGAAC 180

Qy 193 gacttgaggtaaagatcccttggagagtggtgtgtcggcgagcttctgt 252

Db 181 GACTTGGAGGTAAAGGATCCCTTTGGAACGTGTGTGTGCGCCGCTCTCTGT 240

Qy 253 ctgagtgctgttcttgatgtgagccttgggttgacgctgtctctccaaacgttaa 312

Db 241 CTGAGTGTCTGT 300

Qy 313 ggaactgagagactgtgacagacgtgtctaaagagatcacagcctgtccacccttggg 372

Db 301 GGACTGGAGAGCTGTGATACAGACAGCTGTAGAGAGATACAGGCTGCCACCTGGGG 360

Qy 373 acgccccggagtgagagagagccagagagcctgtgtctctcactcgtcgtcagag 432

Db 361 ACGCCCCGGAGGTGGGAGAGCCAGGAGCGCTGTGGTCTCTACTGTGGTACAGAG 420

Qy 433 accgagtgctgttgaagcgaagcttccctcccgccgagccttcttggct 492

Db 421 ACCGAGTCTGT 480

Qy 493 gcttggagagcgagcgagcgtgtgtgtctgtgactgtgtgttctgtctgtgtgt 552

Db 481 GCTTGTGAGAGAGCGGAGCGGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540

Qy 553 tcttctgt 612

Db 541 TCTTGT 600

Qy 613 ttagtctgactcgcacatctgactgaagttagatccagagcctcactaattgtgacttc 672

Db 601 TTAGTTGATCTCGACATGTGACTGAATGTATCCAGGCTCTAATTTGTCTAGTTC 660

Qy 673 aggttaagagagagacttggcagacttctgtgtccttgaatgagccacatcgtgtgt 732

Db 661 AGGTAAAGAGGAGCCTTGGAGACTTCTGTGCTCTGAATGGCCAAATTCGATGTGT 720

Qy 733 gatggcaatcagagagagacttaattctgaataatcctgtgtgtgtgtgtgtgtgtgt 792

Db 721 GATGGCAATCAGAGGAGGACTTTAATTTGAATTAATCTGTGCTGTGAATTCATTTT 780

Qy 793 ttcagactgagacccgactcactcactgaagagccactatactcactgaagtgagagatt 852

Db 781 TTCAGACTGAGACCGGCTCTTATCTCGATCAGAGAGCCCTATCTTACGTGTGCAAGATT 840

Qy 853 tggcagaagatcctccgcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 912

Db 841 TGGCAGAAATCTCTCGCCTGATGTTAAACCATGCTAAATTAACCAAGAAAGCAAGGTC 900

Qy 913 cccgaatcctgtctcttggagagaaacaaacactgtgtgtgtgtgtgtgtgtgtgtgtgt 972

Db 901 CCGAATCTGTGCTCTTGGAGAGAAACAAACACTGTGCGGCAAAAGTGTGAGCCCTCTT 960

Qy 973 cctgtatctaccccgagatcagagagccgacgactgtgtgtgtgtgtgtgtgtgtgtgtgt 1032

Db 961 CCTGTATCTACCCGAGATGAGAGAGCGCGGACTTGTGCGGAAACCCCAACTGTGTCTCC 1020

Qy 1033 ccaaccccttaccagacagaggt 1092

Db 1021 CCACCCCTTATCCAGCAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080

Qy 1093 tgggtgagagactgt 1152

Db 1081 TGGTGAAGAGACTGT 1140

Qy 1153 acgagatcgatatacagctgt 1212

Db 1141 ACGAATATGCGATATTACCGCTGCGCACTATGTGCTCTCCATGTCCAGGGGCAATTTGC 1200

Qy 1213 agcccttcaglatgt 1272

Db 1201 AGCCCTTCAGATATTGGCTTTTCTTCTGTGAGATCTCTAATTTGAGAAATCAACATC 1260

Qy 1273 ccccttctcagagagatcccaagcctcagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1332

Db 1261 CCCCTTCTCGAGAGATCCCAAGCCTCAAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320

Db	3479	CGTAAACCAAAAAAGGGGATTCCTCTGGGCTCCAGACACAGAGGACGATTTGATG	3538
QY	3553	ctatcaaaaaggccctgctgagcgacactgctgcccctccctgacgtatacaaccct	3612
Db	3539	CTATCAAAAAGGCGCTGCTGAGCGCAGCTGCTGGCCCTCCCTGACGTAACTAAACCT	3598
QY	3613	ttacccttatgtgtagtagcgtaagagtagaccgaggaagtttaaccnaaaccttag	3672
Db	3599	TTACCTCTTTATGTGGATGAGGTAAAGGAGTAGCCGAGGAGTTTAAACCAACCTTAG	3658
QY	3673	gaaccaaggagagacactgtgcttaaccgttaaaaagacttgatccctgtatagcaagt	3732
Db	3659	GACCATGAGAGGAGACCTGTTGCTACCTGTAAGAGCTTGATCCCTGTAAGCCAGTGCTT	3718
QY	3733	ggcccgatgtctgtagaggtactgacagctgtgcccatactgttcaagagacgtcgacaat	3792
Db	3719	GGCCCGTATGTCTGAAGGCTATCCGACCTGGGCTACGTGTCGTCAGAGAGCGCTGACAAAT	3778
QY	3793	tgactttggagacagataaactagtgtaagcccccaatgcatgtagaagacatgctgagc	3852
Db	3779	TGACTTTGGAGACAAATATTAATCTTAATAGCCGCCCATGCAATTTGAGAACTGCTCGGC	3838
QY	3853	agcccccagacagctgtagtagcaaacgcccgaatgaaccataatcaaaagctgcttcta	3912
Db	3839	AGCCCCAGACCCGATGATGATGACCAAGCCCGCATGACCCATTCATAAGCTGCTCTGA	3898
QY	3913	cagagaggtcacttctgctccacagccgcttcaacccctgcaactcttctgcttag	3972
Db	3899	CAGAGAGGTCACCTTTCGCTCCACAGCCGCTCTCAACCTGCTCACTCTTCTGCTGAAG	3958
QY	3973	agactgttagaacagtagactcaatgactgcaataatgtagtagaagagactgtagtcc	4032
Db	3959	AGACTGATGAAACAGTAGACTGATGATTTGCCATCACTATTGTTAGGAGACTGGGGTCC	4018
QY	4033	gcaagagaccttaagacataacgcgtgactgtagaagtgtcaactgtgtcaactgtagaa	4092
Db	4019	GCAAGGACCTTACAGACATACCGCTGACTGAGAGAGTGTACTGTAACCTGTTCACTAGCGAA	4078
QY	4093	gcaactatgtgtgtagaagttaagagatgagctgggagcgctgtgtgtagcgagaccca	4152
Db	4079	GCAGCTATGTGGTAGAGTAAAGAGATGAGTGGGGGGCGAGTGTGACCGGACCGCA	4138
QY	4153	cgactgggagcagcctgcccgaagaagactcaagcgaagaagctgtagcttagtccc	4212
Db	4139	CGATCTGGGGCGAGCAGCCGCGGAAGGAACTCAAGCCCAAAAGGCTGAGCTCAAGGCC	4198
QY	4213	taacgaacacttgagctgagccgaaggaatacctaataacattatagagacagcagt	4272
Db	4199	TCACGCAAGCTTTGGGCTGGCGCGAAGGAAATCCATTAACATTTATACGAGACAGGT	4258
QY	4273	atgctctgagctgacagctacacgtgagcggtcattaaacaaagggtgtgctaact	4332
Db	4259	ATGCTTTGGGCTGACACAGTACACGGGGCATCTTTAAACAAAGGGGTTCTTTAACTT	4318
QY	4333	cagcagggaggaataaagaacaagaagaaattctaaagcattatagagccttcaat	4392
Db	4319	CAGCAGGAGGGAATAAAGAAAGAGAAATTTAAACCTATTAGAAAGCTTTACATT	4378
QY	4393	tgccaaaaggctagctattatatactgctgtagacatcaagaagccaaagatctaat	4452
Db	4379	TGCCAAAAGGCTAGCTATTATATACACTGCTCTGACATCAAGAAAGGATCTCATAT	4438
QY	4453	ctaaagggagacagatgctgtagccggtgtgcaagcagagcagcagctgtttaaacttc	4512
Db	4439	CTAAGAGGAAACAGATGGCTGACCGGGTTCCAAAGCAGGAGCGCCAGGCTGTTAACTTC	4498
QY	4513	tgctataatagaacgccccaaagccccagaaacccagacgacagtaacacctagaagct	4572
Db	4499	TGCTTATATATGAACGCCCAAGGCCCAAGAACCCAGACAGACAGTACACCTTAGAAGCT	4558
QY	4573	ggcaagagataaaaagatagacagttctctgtagactcggaggggactgtctatact	4632

Db	4559	GGCAGACATTAATAAAGATAGACAGTCTCTGAGACTCCGAGGGGACCTGCTATACCT	4618
QY	4633	catatgggaaggaatactctgcccacaaagaggttagaataatgattccaaagatacatc	4692
Db	4619	CATATGGGAGGAATTCCTGCCCAACAAGAGGTTTAGATATGTTCAACAGATATCATC	4678
QY	4693	gtctaacccaccttaggaactaaacaactgacagctgtgtcagaataatcccttataatg	4752
Db	4679	GTTTACCCACCTAGGAACCTTAACACCTGCAGCGAGTGTGTGACAGATCCCTTATCATG	4738
QY	4753	ttcttaaggtacccagaggtgagctgactcgggtgtgttcaaaactgtgtgcccctgacgtg	4812
Db	4739	TTCTGAGGCTACAGAGACTGCTACTCTGGTGGTCAAACTTTGTGTGCTCCCTGCCACTGG	4798
QY	4813	ttaatgctaatcccttccagaataactcccaagaaagagactcaagggagagcccccaggg	4872
Db	4799	TTAATGCTAATCCTTCAGAAATACCTCCAGAAAGAGACTAAGGGAGACCCACAGGGCG	4858
QY	4873	ctcaactgggaagttagacttaactgaggttaagccggcctaatacggnaaaataatctat	4932
Db	4859	CTCACTGGGAAGTGAAGTCTGAGAGTAAAGCCGCTAAATACGGAACCAATATCTAT	4918
QY	4933	tggtttttagaacaaccttccagaatggtagaggtctatactctaaagaagagactt	4992
Db	4919	TGTTTTTTGTAGACACTTTTCAGATGGGTAGAGGCTTATCTCTAAGAAAGAGACTT	4978
QY	4993	caaccggtgtgtgcttaagaaataactgtaggaataatttccaaagtltgaaatacctaag	5052
Db	4979	CAACCGTGGGTGTAGAAATACTGAGAAATTTTTCAGATTTTGAATTAAGTAAAG	5038
QY	5053	taataaggttagaacaatgcttccagcttctgttcccaagtagtaagagactgtgccaaga	5112
Db	5039	TAAATAGGCTAGACATATGCTTCCAGCTTTGTCGCCAGGTAAAGTCAAGGACTGGCAAGA	5098
QY	5113	tattggagattgatttggaaactgcatgtgtcaacagcccaaaagctcagagacagtag	5172
Db	5099	TATTTGGGATTTATTTGAAATGCTATGTCATGACAGCCCAAAAGCTCAGAGACAGTAG	5158
QY	5173	agaagatgtaataagaaccttaagaagaccttaacaaattgaacacagagagactgcatla	5232
Db	5159	AGAGATGATTAATGAACCATTAAGAGACCTTACCAAAATTTACACAGAGACTGCAATTA	5218
QY	5233	atgattgtagtgccttccctgccccttgtgctttttaggttagaagaacccctggaagct	5292
Db	5219	ATGATTTGATGCTGCTCTCCGCTTTGTGCTTTTAGGGTAGAGAACCCCTTGACAGT	5278
QY	5293	ttgggtcgaacccctcagtaaatgtctctacggyggagcccccctgtgtagaagaattgct	5352
Db	5279	TTGGGCTGAACCCCTATTAATTTGCTTAACGGGGAGCCCCCGTTGGAGAAATTTGCT	5338
QY	5353	ttgcacatagtcgtatgtgtgtcttccagacctgttctcttaggtcaagcgctcg	5412
Db	5339	TTGCACATATAGTGTGATGTGCTGCTTTCCAGGCTTTGTCTTAGGCTCAAGGCGCTCG	5398
QY	5413	agttggttagagcagcagcgltggaagcagctccggagagcctactcaggagagagacttcg	5472
Db	5399	AGTGGTGAGGACAGCAGAGCGCTGGAAGCAGCTCCGGGAGGCTACTACAGAGAGACTTCC	5458
QY	5473	aagttccaactcgccttccaaggttggagattcagctatgtttagaagcagcgtlgaagaa	5532
Db	5459	AAAGTTCACATCGCTTCCAAAGTTGAGATTAATGATTAAGACCCACCCCTGACAGAA	5518
QY	5533	acctgagactcgggtggaagagacattatctgtaacttttgacacacaaagaggtgtga	5592
Db	5519	ACCTCGAGACTCGGTGGAAGGAGCTTATCTGTGACTTTTGAACACACCAAGGCTGTGA	5578
QY	5593	aagtgcgaagaaatcccaactgtgataatccatgcccaogtlaagccggcgccactcccg	5652
Db	5579	AAAGTGAAGGAATCCCACTGATCCATGATCCACAGCTTAAGCGGCGCACACTCCCG	5638
QY	5653	actcgggttagaagcagaaagagactggaatacccttaagtttgcctcatcagctcgg	5712
Db	5639	ACTGGGGTGGAGAGCGCAAAAGACTGAGATTCCTTTAAGCTTGCCTCATGCGCTGG	5698

QY 121 tttcttttacttgcccttcaatttggttgccgggaaatcctgcgaccaccctta 180
Db 687 TGTTCCTTTACTGGCCCTTTCATTTGGCTTGCCGGGAAATCTCGCACACCCCTTA 746
QY 181 caaccggaacccgacttgagaglaaaggatccctcttggaacgtgtgtgtgcgc 240
Db 747 CACCCGGAACCGACTTGAGAGTAAAGGATCCCTTTGAAAGTGTGTGTGTGTGCGGC 806
QY 241 ggcgtctctgtcttgagtgctgttctcggttgatgcgcgtcttggttgagcgtgct 300
Db 807 GGCCTCTCTGTTCGAGTGTCTGTTCGGGTGATGCGCCTTTCGGTGTGCACTGTCT 866
QY 301 ctgaagccctgaagacttgagagacttgatcaagaacgtgtctaggaagatcaagagctg 360
Db 867 CTCAGACCGTAAAGACTTGAGAGACTGTGATCAGACAGCTGTAGAGAGATCAAGAGCTG 926
QY 361 caacccttgaggagcccccggagggtggggagagccaggaacgctgtgtgtctctact 420
Db 927 CCACCTTGCGGAGCGCCCGGAGGTGGGAGAGCCAGGAGCGCTGTGTGTCTCTACT 986
QY 421 gtcggtcaagagaccgagttctgttgttgaaagaaagcttccctcccgccgctgcg 480
Db 987 GTTCGGTCAAGAGACCGAATTCGTGTGTGAAGCGAAAGCTTCCCTCCGCGCGCTCG 1046
QY 481 actctttgcctgtcttgagaaagcggagagctgcgtgtgtctgagatcttggtt 540
Db 1047 ACTCTTTTGCTCTGCTTGAGAGAGAGCGGAGGCTGTGTGTGTGTGTGTGTGTGT 1106
QY 541 ctgtctctgtgtctgttctgtgtgcctgtctgtctcaagtttaataatgagacagag 600
Db 1107 CTGTTTCGTGTCTTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1166
QY 601 tgaactaaccctttagttgactctcgacacttgagactgaagtagatccagagctcata 660
Db 1167 TACAGACCCCTCTGATTTAACTCTGACACATGGAAGCTTGAAGTTAGATCCAGGCTATA 1226
QY 661 attgtgcaattgaaglaaagaaagacttgacagacttctgtgcctctgaaatgagca 720
Db 1227 ATTTGTCACTTGAAGTAAAGAGGAGACTTGAGACACTTGTGAGCTTAATATGAGCA 1286
QY 721 catcgatgttgatgagccatcaagagagacctaattctgaaatatactgtgtgt 780
Db 1287 CATTCATGTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1346
QY 781 aggaactcttttcaagactgagccggtctcatcctgatacaagagccatctctta 840
Db 1347 AGGCAATCATTTTTCAGACTGAGCCAGCTCTCATCTGATCAGAGAGCCCTATATCTTA 1406
QY 841 cgtggaagaattgagagaagatcctccgcaatgggttaaacatlgcttaataaaca 900
Db 1407 CGTGGCAAGACTTGAGAGAGATCCCTCGCCCATGAGGTTAAACATGAGTTAAACAA 1466
QY 901 gaaagccaggtlcccgaaatcctgtgtcttgagagagaaacaacaactcgcgcgaagaag 960
Db 1467 GAAAGCCAGGTCCCGAATCTGTGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1526
QY 961 tgaagcccttctctgtatatacccgagatcgagagacggccgactgtgcgaagacc 1020
Db 1527 TCGAGCCCTC-TGCTGTATCTAACCCGAGATGAGAGAGCCCGCACTTGGCGGAGACC 1585
QY 1021 caactgttccccaaccccccttaaccagacagaggtgtgtgtgagagga-cctctgtccct 1079
Db 1586 CAACCTTGTCCCAACCCCTTATCCAGACAGAGGTGCTGTGAGGAGACCTCTGCGCT 1645
QY 1080 ccttgagatccggtlgtgagaggaactgtctgcggagactcgaagccgagagagcgacac 1139
Db 1646 CCTGAGACTCCGGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705
QY 1140 ccgagagcgagacagagatcgagatatatacgctgcaacatagagccctccatgcca 1199
Db 1706 CCGAGCGGAGACAGAGAGATCGGATTTACCGCTGGCACTATATGAGCCCTCCAGAGCCG 1765
QY 1200 gggggccaattgagcccccccaaglatatgaccttctctgtcgaatctcataatgtg 1259

Db 1766 GGGGGCAATTTGAGCCCTCCAGATATGAGCCCTTTCTTGTGAGATCTATATATGG 1825
QY 1260 aaactaacatccccccttctcgagagatcccccaacgctcaaggggttgatgagctc 1319
Db 1826 AAACCTAACCATCCCTTTCTCGAGAGATCCCAAGGCTCACGGGAGTGTGTGAGAGTTC 1885
QY 1320 ctatgttctcaacagactacttggatgatttgcacaagctgtctgagacactctc 1379
Db 1886 CTATGTCTCTCTACAGGCTACTTGGATGATTTGCAACAGCTGTGAGACACTCTTC 1945
QY 1380 acaaccgagagcgagagagaatctcttgagagcttagaanaaaatgttctctggggcgagc 1439
Db 1946 ACAACCGAGAGAGAGAGATTTCTTTAGAGGCTAAGAAAATATGCTTCGGGCGGAGC 2005
QY 1440 ggcgaaccacagcagttgcaaaatlgatatgagatcttcccttgactgcgccccgt 1499
Db 2006 GGGGAGCCACAGAGTTGCAAAATGAGATTTGACATGGGATTTCCCTTGACCTCGCCGCT 2065
QY 1500 tggagatacaacagcttgaaagtgaagagagcttgaaatctatccgacagctctgtgtg 1559
Db 2066 TGGGACTCAACACGGCTGAAGGTAGGAGAGCTTGAAATCTATCCCGAGGCTGTGTG 2125
QY 1560 ggcgttctcgggcgccctcaagacgcccactaatttgctaaagtgaagagatgagt 1619
Db 2126 GCGGCTCTCCGGGCGCTCAAGAGCGCCCACTAATTTGGCTAAGTAAAGAGAGGTGATG 2185
QY 1620 caggagaccgaacactccctcgatattcttgagagctcatggaacctcgaagcgg 1679
Db 2186 CAGGAGCCGAAGCAACCTCCCTCGGTATTTCTTGAAGAGGCTCAAGAGGCTTCAGCGG 2245
QY 1680 ttcaccccttctgactcactcaagagagccagaagaacctcgaagagccgagccatc 1739
Db 2246 TTCACCCCTTTTGATCTTACCTCGAGGCGCCAGAGAGCCTTAGTGCCCTTGCTTAT 2305
QY 1740 gggcagctgcgtctgatalatacagaagaactctgaagacttgaaaggtttacagagagct 1799
Db 2306 GGGCAGTGGCTCTGTGATATCAGAAAGAACTTCAAGAGCTGAAAGGTTTACAGAGAGCT 2365
QY 1800 gagttacgtatctagtgagagagcagaagaagtgtatctcaagaagggagacagaagag 1859
Db 2366 GAGTTACGTGATCTAGTGAAGAGGCGCAGAGAGGAGTGTTCAGAGAGGAGACAGAGAG 2425
QY 1860 gagaaggaacagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1919
Db 2426 GAGAGAGAACAGAGAAAG 2485
QY 1920 caagagaagaatttgactaagaatctgtgcgaggtgttgaaaggaagcagcagagagag 1979
Db 2486 CAGAGAGAAATTTGACTAAGATCTTGGCGCAGTGTGAAAGGAGAGCAGCAGAGGAG 2545
QY 1980 agagagagagattttagaanaattaggtcagccctagaacagctcagaggaactcggagat 2039
Db 2546 AGAGAGAGAGATTTTAGGAAATTTAGGTGAGGCGCTTAGACATCTGAGAGAGCTGGGCAT 2605
QY 2040 aggaacccactcgacaaaggaacaggtgtgtgttgtaaagaagaagcaactcggcaag 2099
Db 2606 AGGAGCCCACTCGACAAAGGACAGTGTGTGATTTGTAAGAGAAAGAGACACTGGCAAG 2665
QY 2100 aactgcccccaagaaggaacaaagaacaggaagctcagactcctaagaagaataaagat 2159
Db 2666 AACTGCCCAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2725
QY 2160 taaggagacggggttcgagccccctcccgagccagagtaactttgaaggtgagaggg 2219
Db 2726 TAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2785
QY 2220 caacagttgagttcctgtgtgtatccgagagcagagacatctgaagctgtctgaacacat 2279
Db 2786 CAACCACTTGAAGTCTCGTTGATACCGGAGCGAGACATTCAGTGTGCTTAACACATTA 2845
QY 2280 ggaataactaaagaanaaaatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2339

Db 2846 GGAAGCTAAAGAAAAAATCTGGGTGATGGGTGCGACAGGCGCAAGCGAGTATCA 2905
Qy 2340 tggactaccgaagaaccgttgaacttggagtggaacgggtaaccactggtttcgtc 2399
Db 2906 TGGACTACCCGAGAAACGGTTGACTTGGCACTGGAGCGGGTAACCCCTGTTCTGTGTC 2965
Qy 2400 atcccttggtgcccgaatacccttctagtagtagaacttactgaacaaatggagatcaa 2459
Db 2966 ATCCCTGAGTGCACAGTACCCCTTCTAGTAGAGACTTACTGACCAAGATGGAGGCTCA 3025
Qy 2460 attcttctgaacaagaagaccgaagtgcttctgaatacaaaaccatctgtgttg 2519
Db 3026 ATTCTTTTGAACAAGAGAACGAGAGTGTCTGTGATATACAAACCCATCAGTGTGTG 3085
Qy 2520 accctcaatagatgatgaalacgaactaattctcccaagtaaacgctgaatcaat 2579
Db 3086 ACCCTCCAAATTAGATGATGATATCGACTATATCTCTCCCAAGTAAAGCTGATCAAGAT 3145
Qy 2580 atacagctcgtgttgagagagtttccccaagccttgagcagaacacgaggaatggattg 2639
Db 3146 ATACAGTCTGTGTGGAGAGATTTCCCAAGCTGGGCAAAACCCAGGAGATGGGTTTG 3205
Qy 2640 gcaagaagagttccccaagtgatlatcaactgaagcagtgctacacagatlatcagtc 2699
Db 3206 GCAAGGCAAGTCCCGCACAGGTTATTCACACTGAGGCCAGTCTACACAGTATCAGTC 3265
Qy 2700 agaaatgaccccttgagtagaagagctcgagaagaatttggccgcatgttcaaatga 2759
Db 3266 AGACAGTACCCCTTGAGTATGAGAGGCTCGAAGAGAAATTGGCCGATGTCAAGATTA 3325
Qy 2760 atccaagaagcattcctagttcctgtccaatcccttgaaatcctccgttaccggtt 2819
Db 3326 ATCCACAGGGCATTCAGTCTCTGTGTCATCCCTTGGAAATATCCTCTGTACCGGTT 3385
Qy 2820 aggaagccttggagccaatgattatcgacaagacttgaagagagtcataaag 2879
Db 3386 AGGAAGCCTGGACCAATGATTCAGACAGTACAGAGACTTGAGAGAGTCAATTAAGG 3445
Qy 2880 gtgcagagacatacaccacaaggtcccgaaaccccttaacactttagcgccctccgct 2939
Db 3446 GTGCAGAGACATACACCAACGGTCCGAAACCTTATTAACCTTTAGGCGCCCTCCGCT 3505
Qy 2940 gaaaggaactgtacacagatatttgaactaaagaatgccttcttgccttagattac 2999
Db 3506 GAACGGAGACTGGTACACATATTGACTTAAGATGAGATCCAGGTACCGGAACCGGG 3565
Qy 3000 cccactagccaacacttttgccttcgaatggagaagatccaagtaacggagaacggg 3059
Db 3566 CCCACTAGCCAAACGCTTTTGCCTTCGATGAGAGATCCAGGTACCGGAACCGGG 3625
Qy 3060 caagtcaccttgaaccgaactgcccgaaggttcaagaactcccgacacatcttgaaga 3119
Db 3626 CAGCTACCTGGACCCGAGTCCCAAGGTTCAAGAACTCCCGACCATTTTGAAGAA 3685
Qy 3120 ggccttaacagaagacttggccaacttcaagatccaacacccatgaagtacccctcag 3179
Db 3686 GCCCTACACAGGAGCTGGCCAACTTCAGATGCCAACACCTCTAGGTGACCTCTCTCAG 3745
Qy 3180 tacgttgaatgacttcttgcgcggagccacaacagagacttcttagaagatgaag 3239
Db 3746 TACGTGATGACCTGCTTGTGGCGGAGACCAACAGAGACTCTTGAAGAGGTACAG 3805
Qy 3240 gaactactgttgaatgttctgaacctagctacagagccttctgtaagaagcccgaat 3299
Db 3806 GCACACTGCTGGAATTTCTGACCTAGGCTACAGAGCTCTGTGAAGAGGCCCAAT 3865
Qy 3300 tgcagaagaagatgaacttgggttgaactgttcgaggcgagcagaatgactgaag 3359
Db 3866 TGCAGAGAGAGGTACACTTGGGGTACAGTTTGGGGGCGGAGCGATGGCTGAGC 3925
Qy 3360 gagagcaggaagaactgtagtccagataccggcccacaacacagcaagaagtgaga 3419
Db 3926 GAGGACGGAAGAAACTGTAGTCCAGATACCGGCCCAACACAGCAAGCAAGTGA 3985
Qy 3420 gaatttttggagacagctggaatttgcagactgtgatacccggttctgcacttaaga 3479
Db 3986 GAGTTTGGGAGACAGCTGATTTTGCAGACTGTGATGCCGGGTTTGGACCTTAGCA 4045
Qy 3480 gcccaactctaccgcttaaccaagaagaagggagatctcccttgggtctctgaacag 3539
Db 4046 GCCCAGCTCTACCCGCTTACCAAGAAAGAGGGAATTCCTGGGCTCTGTGACACAG 4105
Qy 3540 aagcatttgaatgatacaaaagccctgtagcagcagctgtcttggccctccctgac 3599
Db 4106 AAGCATTTGATGTATCAAAAAAGGCCCTGTGAGGCGACACTCTGTGGCCCTGTGAC 4165
Qy 3600 gtaactaaaccccttacccttatttggatgtagcgtgaagggatagccgagagttta 3659
Db 4166 GTAACCTAAACCTTTTACCTTTTATGTGATGAGACCTGAAGGAGATGACCCGAGAGATT 4225
Qy 3660 acccaaacccctlagagacatgagagagacccgtgtgtcctactcttcaagaagcttga 3719
Db 4226 ACCCAAAACCTAGAGACCATGAGAGACCTGTGCTTACCTGTCAAGAGCTTATGCT 4285
Qy 3720 gtagcaggtgttggccgtagtctgaaggtatcgcagctgttggccaacttgg tcaag 3779
Db 4286 GTAGCCAGTGTGGCTGGCCGTATGCTGAAGGCTATGCGAGCTGTGGCATATGCTGTA 4345
Qy 3780 gaagctgaacaattgacttggagagagataataactgtataagcccccaatgag 3839
Db 4346 GAGCGTACAAATTTGACTTTGGAGCAAGATATACATGTAATAGCCCCCATGCAATTTG 4405
Qy 3840 acaatgcttgcagcccccagacagatgatatgaacaaagcccgatgacccaatcaa 3899
Db 4406 AAGATGTTGGGAGGCCCCAGACCGATGATGATGACCAAGCCCGATGACCATATGA 4465
Qy 3900 agccttcttcaagaaggggttcaacttctgttccaaacagccgcttcaaccctgca 3959
Db 4466 AGCCTGTTCTCAAGAGAGGCTCACTTTCCTCCACAGCGGCTTCACCTGCGCACT 4525
Qy 3960 ctcttgcctgaagaagactatgaacagtgatcaatgttccatcaactttagatgag 4019
Db 4526 CTTCTGCTGAAGAGACATGATGACAGTCAATGATTGCTATCAATTAATTTGAG 4585
Qy 4020 ggaactgggttccgcaagaaccttaacagacataccgctgaagtagaagtgtaacctg 4079
Db 4586 GAGACTGGGGTCCGAAAGCACTTACAGACATACCGCTGAGTAGAAGTCTTAACCTG 4645
Qy 4080 ttaactgaaggaagcagctatgttggtaaggtlaagagatgggttggcgagtggtg 4139
Db 4646 TTCACGTGAGAGAGACAGCTATGTGTGTAAGGTATAGAGATGCTGGGGCGCGGTG 4705
Qy 4140 gaaggaacccgagacatcttggccagcagccttgcggaaggaattcagcgcaaaagct 4199
Db 4706 GACGGAGCCACACGATGTGGGCGAGCCTGCGGAGAGAACTTCAGGGCAAAAGCT 4765
Qy 4200 gaacttgaagccctgaagcaagcttgggttgcgcgaaggaataccataaacttat 4259
Db 4766 GAGCTCATGCGCTTACCAAGCTTGTGGCTGCGCAAGGGAATTCATTAACATTAT 4825
Qy 4260 acggaacaggtatgaccttgcgacttgcacagctacacgggggcatctataaacaag 4319
Db 4826 ACGGACACAGGTATGCTTGGCAGTGCACAGCTACACGGGGGCACTTAAACAAAG 4885
Qy 4320 ggttgttacttaagcagggagggaaataaagaacaagaagaaattcgaactatla 4379
Db 4886 GGGTGTCTTACTCAGCGAGGGAATTAAGAAACAAAGAGAAATTTAAACCTATT 4945
Qy 4380 gaagccttaacttggccaaaagctagctattataaactgtcttgaatcagaagcc 4439
Db 4946 GAAGCTTACATTTGCCAAAAAGCTGTCTATTTATACGTCTTGACATCGAAGAGCC 5005
Qy 4440 aaagctcatalatagagggaacagatgcttgaacgggttcccaagcagcgagccag 4499
Db 5006 AAAGATCTATATCTAGAGGGAACAGATGGCTGACCGGGTTGCCAAGCAGGAGCCAG 5065

QY 4500 gctgttaacctctgtcctaataaagaagcccaagccccaagccccaagcagctaac 4559
|||||
Db 5066 GCTGTTAACTTCTGCTTAAATAGAAAGCCCAAGCCCAAGACCGACGACAGTAC 5125
QY 4560 accctagaagacttgcaagagataaagaatagacccgttctcgaaccccggaagg 4619
|||||
Db 5126 ACCCTAGAAAGACTGCGCAAGAGATAAAAAAGATAGACCAATCTCTGACACTCCGAGGGG 5185
QY 4620 acctgtataacctcataitggaaggaatcctgcccacaagaagggttagaatatg 4679
|||||
Db 5186 ACCGTGATACCTCATATGGAAGAAATCCTGCCCAAGAAAGGGTTAGAAATATGTC 5245
QY 4680 caacagatatcgtcttaaccacctaaggaactaaacacctgacagatgttgtaaga 4739
|||||
Db 5246 CAACAGATACATGCTGCACCCACCTAGGAACATAAACACCTGACAGTGTGTCAGAAC 5305
QY 4740 tcccctatacattctcgaagcttaacagagatggtgactgctgtgtaaatctgtg 4799
|||||
Db 5306 TCCCTTATCATGTTCTGAGGCTACCGAGTGGTACTCGGTGTCATAACATTTGTG 5365
QY 4800 cccgtgacagctgtgttaatgttaactcctccagaatacctccagaagaagactaagg 4859
|||||
Db 5366 CCGTCCACAGCTGTTAATGCTAATCCTCCAGAAATACCTCCAGGAAGAGACTAAGGG 5425
QY 4860 agccaccacagcgtcactggtgaagatgacttaactgaggttaagccggttaataag 4919
|||||
Db 5426 AGCCACCCAGGCGCTCACTGGAGAGTGACTTCAAGGATTAAGCGGCTAATATCG 5485
QY 4920 aacaaatactattgtgtttgttagacacattcaagaaatggtgtaagggttaact 4979
|||||
Db 5486 AACAAATATCTATTGTTTGTGACACACTTTTCAGGATGGGAGAGGCTTATCTACT 5545
QY 4980 aagaagaagactcaaccgtgtgtgtgaagaataactgtaagaaatlttccaagatt 5039
|||||
Db 5546 AAGAAAGAGACTTCAACCGTGTGGCTAAATAAATACTGAGGAAATTTTCCGAGATT 5605
QY 5040 ggaataactaaggttaataggttcagaaatggttcacgttctgtgccaagtaagtca 5099
|||||
Db 5606 GGAATTAACCTAAGTATGAGGCTGACAAACATGTCACACTTTTGTGGCCAGGTAAGTC 5665
QY 5100 ggaactgccaagatacttggtggtatgattggaactgcaatgtgtaacagaacccaa 5159
|||||
Db 5666 GGAATGCGCAATATTTGGGGATTGATGGAACATGCAATGTCATACAGCCCAAGC 5725
QY 5160 tcaaggaagtagaagagatgaataagaacatlaaagaccccttaacaaatlgacaca 5219
|||||
Db 5726 TCAGGACAGGTAGAGAGATGAATGAACCAATTAAGAGACCTTACCAATTAACACA 5785
QY 5220 gagactgacataatgattgtagtctcctgccccttgctgtctttaggtgtagga 5279
|||||
Db 5786 GAGACTGGCATTAATGATTTGATAGCTCTCCCTTGTGCTTTTAGGGTTAGGAA 5845
QY 5280 acccttggaagatttggtggtgaacccctatgaattgtctcagggggaaccccccg 5339
|||||
Db 5846 ACCCTGAGACAGTTGGGTGACCCCTATGAATGCTCTACGAGGGGACCCCCCGT 5905
QY 5340 gcaagaatgctcttcaatatgctgtgtgtgctgtcttcccaaccccttgctc 5399
|||||
Db 5906 GTGAAATGCTCTGTATATAGTGTGATGCTGCTGCTTCCCAACCTGTTCTCTAG 5965
QY 5400 ctaagagcgtcagatgggtgtaggcagcagcgtaggaacagctccgggagccactaca 5459
|||||
Db 5966 CTCAAGGCGCTGAGTGGGAGGCAACGAGCTGGAACAGCTCCGGAGGCTTACATCA 6025
QY 5460 gg--aggaactgtcaagttcacatgcttcccaagttgagatcaagttcatalgt 5516
|||||
Db 6026 GGAGAAAGAGACTTGAAGTTCCACATCGCTTCCAGTGGGAGATTCACATCTATGTA 6085
QY 5517 cgcacacgtgcaagaaaccccgagactcgtgtgaaggaacccatctcgtactttgac 5576
|||||
Db 6086 CGCACCGCTGAGGAACCTCGAGACTCGGTGGAAGGCGCTTATCTCTACTTTTGA 6145
QY 5577 acaccaacgctgtgaagatcgaaagaatcccaactgtagtcaatgcatccacgttag 5636
|||||
Db 6146 ACACCAAGCGCTGTGAAGATCGAAGAAATCTCCACTGATCTCATATCCAGCTTAAG 6205
QY 5637 ccggcgacacctcccgactcggggtgtggaagccgaagaagactgaaatcccttaag 5696
|||||
Db 6206 CTGGCCGCACTCCGACCTCGGGGTGGAGGCCGAAGAACCTGAGAAATCCCTTAAGCT 6265
QY 5697 cgcctcacaatgctgtgttcttactacttaacaaatactcccaagcagtagtaacgct 5756
|||||
Db 6266 CGCTCTCATCGCTGGTTCTTACTATCAATTAATCTCCCGAGCCAGTAGTAACCGCT 6325
QY 5757 tatagaagctggaacccccaatagaccttalcctcttcgtgctgcatattgacccctga 5816
|||||
Db 6326 TATAGACAGCTCGAACCACCCCATAGACCTTATCTCTTACTGCTGATTTATGACCTGA 6385
QY 5817 taagggtgtacagtgaataatagacactgaggtgtgtccttaagggaacccctgtag 5876
|||||
Db 6386 TACGGGTGACAGTGAATATGCACTGAGGTGTGCTCTAGAGGCACTGCTGACCTGA 6445
QY 5877 actgacttctgctccgatltgataaccgctgtlaaagaacacccccaactagt 5936
|||||
Db 6446 ACTGCATTTCTGCCTCCGATTGATTAAACCCGCTTTAAAGCACCTCCACACTAGT 6505
QY 5937 ccgtagttaaggttctatgtctgcccaggaacagagaagaagaataactgtggtgtc 5996
|||||
Db 6506 CCGTAGTATGGGTTCTATTGCTGCCCCAGGACAGAGAAAGAAATACGTGGGATTC 6565
QY 5997 tgggaatcctctgttagaagatggaactggtgtcactccacagatgtagactggaatg 6056
|||||
Db 6566 TGGGAATACCTTCTGTAGGAGATGAGCTGCTGACCTCCACAGATGAGAACGAAATG 6625
QY 6057 gccgactctctccaagacgggttaaatctccttgtgaatccgcccgggcaagt 6116
|||||
Db 6626 GCCGATCTCTCTCCAGGACCGGGTAAATTTCTTTGTAAATTCGGCCGGGCAAGTA 6685
QY 6117 caaataatgaactataataagaataagactgtcccaatcagacttagattataa 6176
|||||
Db 6686 CAAAGTATGAACATATTAAGATTAAGAGCTGCTCCCATACAGACTTAATATCTAAA 6745
QY 6177 gataagtttcaatg--aaaggaacacaggaataatcctcaaaatggtgtaattgtatg 6234
|||||
Db 6746 GATTAATGTTCACTGAAAGAAAGAAACAGGAAATTTCAAAAGTGATTAATGTTATG 6805
QY 6235 ctggggaatagtttttat-tataatggtgggagcaggtgtccacttaacattgct 6293
|||||
Db 6806 CTGGGAATATGTTTATTAATATGCGGGGAGACAGGTCACCTTAACCATTTCCCT 6865
QY 6294 taggataagacagggagacagaacccctgtgtgcaatggtgacccgaataagtcgtgtga 6353
|||||
Db 6866 TAGATAGAGAGAGGGGACAGAACCCCTGTGGAGTGGAGCCGATTAAGTATGCTGA 6925
QY 6354 acaggggccccggccctgtgagacacccgataactgtccggtgcccccaatlaactcgt 6413
|||||
Db 6926 ACAGGGGCCCCCGGCCCTGAGACCCAGCATTAACCTGCGGTGCCCAATTAACCTGCT 6985
QY 6414 gctgactgataacacacagcgcctagacacagatgaactgtatgtatltcctaacacac 6473
|||||
Db 6986 GCGCCCTGACATTAACACACTGCGCTAGCAACGCTGACACTGATGATTTCTTACCAAC 7045
QY 6474 gcttagaactccccaggtgttccgtgttaagaacgaagaagaactcttcacatcaca 6533
|||||
Db 7046 GCTAGAAACTCCCGAGGTGTTCTGTTAAGACAGAGACGAGACCTTCACTCATCA 7105
QY 6534 gggagacttcccaagacataacacacacacacacacacacacacacacacacacacac 6593
|||||
Db 7106 GGGAGCTTTCACACCATCACTCAACGACACCTGATGCTCTTCTTGTGGCTTGG 7165
QY 6594 tctatcctcagggcctccttattatagaaggtatgcttaagaagaagaatcaatgtagac 6653
|||||
Db 7166 TCTATCTCAAGGCGCTTATTAATGAGGGATGCTTAAAGAAAGAAATTAATGTAAC 7225
QY 6654 caaagacatagaataatcaatgtagtgggtgtcccgaaataagcttacccactgaagt 6713
|||||

Db 7226 CAAGAGCATGAATCATGATGATGGGGTCCCGAAATAGCTTACCTCAGTGAAT 7285
 Qy 6714 ttcgggaaggagacatgacataagaaagcccccacatccacacacacttgatag 6773
 Db 7286 TTCCGGGAAGGAGCATGATAGAAAGCTCCCATCCACACACCTTGTGCTATAG 7345
 Qy 6774 tacttggttataagagagcctcagaataatcagatcttacccttgatatacagtg 6833
 Db 7346 TACTGTGTTATAGAGCAGGCTCAGAAATCAATATTGATGCTGTTATAGAGGTG 7405
 Qy 6834 gtggagcatgacacttggtttaaacccttggttccacacttgatcttaaccacaa 6893
 Db 7406 GTGGCAGTGAATGCTGGGTTAACCCCTGTGTTCCACTGAGTCTTAACCAATCCAA 7465
 Qy 6894 agattctggtatgacatggttccacatcgtcccgagtgatctacacatccctgaagagtg 6953
 Db 7466 AGATTTCTGTGCTGATGCTGCAAAATCGCCCGAGTGTACTACCATCTGAGAGAGTGT 7525
 Qy 6954 ccttgatgaaatgactatcgatataccgacacaaagaagacccgtatcccttaacct 7013
 Db 7526 CCTTGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7585
 Qy 7014 agctgtatgctcgtgattaggaagcggcttggtcgttagaagacagagacgtcctgat 7073
 Db 7586 AGCTGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7645
 Qy 7074 cacagagacacagacagacagacagacagacagacagacagacagacagacagacag 7133
 Db 7646 CACAGAGCAC 7105
 Qy 7134 ttcggagccttaaaagagctcgtttagacacacacacacacacacacacacacacacac 7193
 Db 7706 TCTCCAGGCTTATAGAGAGTCTGTTAGACACTAGACAGTCTGATCTTGTGCTGA 7765
 Qy 7194 agtggcttaccagaaacag 7253
 Db 7766 AGTGGTCTACAGAACCGAGGGGATGATGATGATGATGATGATGATGATGATGATGAT 7825
 Qy 7254 tgcagccttaaaag 7313
 Db 7826 TGCACGCTTAAAGAGAGATGTTGCTTATGATGATGATGATGATGATGATGATGATGAT 7885
 Qy 7314 catgacagccttagaagaaagagagagagagagagagagagagagagagagagagag 7373
 Db 7886 CATGAGCAAGCTTAGAGAGAGGTTAGAGAGGCTGAGAGGAGAGAGAGAGAGAGAGAG 7945
 Qy 7374 gtggcttgaag 7433
 Db 7946 GTGTTTGAAGAGATGTTCAACAGTCTCTGATGATGATGATGATGATGATGATGATGAT 8005
 Qy 7434 ggggccccctagtagctctgctctgctctgctctgctctgctctgctctgctctgct 8065
 Db 8006 GGGACCCCTAGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8065
 Qy 7494 tttgctcttgtagaag 7553
 Db 8066 TGTGCTCTTGTAG 8125
 Qy 7554 ccaagagccttcagcagaag 7613
 Db 8126 CCAAGGCTTCTGAGCAG 8185
 Qy 7614 ctattacaag 8245
 Db 8186 CTATTAAACAGACAG 8245
 Qy 7674 acccaggaagtaataaag 7733
 Db 8246 ACCCAGAGAGTATATAAAGCTCTAATAGCCCGAGATTCAGAGACCTGCTGCTGCA 8305
 Qy 7734 gtaaatagtagaag 7793
 Db 8306 GTAATAGTAG 8365

Qy 7794 agatacaggaatgagtagtgcatacgtctatctgattctgttaaacctgactggcacc 7853
 Db 8366 AGATTAACAGGAATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8425
 Qy 7854 atagaagaatgattacacatgacagccctagtgacatctcacaactgcaactgctac 7913
 Db 8426 ATGAGAGATTTGATTACATGATGACAGCCCTAGTACACTATGATGATGATGATGAT 8485
 Qy 7914 tctgccaagagccacagcag 7973
 Db 8486 TCTGCCAG 8545
 Qy 7974 ggaagcggcctcgtatatttaaatgattgtcca 8011
 Db 8546 GGAGCGCGGCTCGATATTATAATGATTGTGTCA 8583

RESULT 4
 LOCUS PEN133817
 DEFINITION Porcine endogenous retrovirus type C proviral gag, pol and env genes and LTR (class A, clone 42).
 ACCESSION AJ133817
 VERSION AJ133817.1 GI:6688947
 KEYWORDS env gene; env protein; gag gene; gag protein; pol gene; pol protein.
 SOURCE porcine endogenous retrovirus.
 ORGANISM Porcine endogenous retrovirus.
 REFERENCE 1 (bases 1 to 8849)
 AUTHORS Toenjes, R.R.
 TITLE Molecular Characterization of Human-tropic and Replication-competent Porcine Endogenous Retroviruses 2 (bases 1 to 8849)
 JOURNAL Unpublished
 REFERENCE Toenjes, R.R.
 AUTHORS Toenjes, R.R.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology, Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225 Langen, GERMANY
 FEATURES
 source
 LTR
 misc_feature
 primer_bind
 gene
 mat_peptide
 CDS
 /organism="porcine endogenous retrovirus"
 /strain="type C"
 /db_xref="taxon:61673"
 /clone="42"
 /note="grown in PK15 (ATCC No. CCL-33) cell line"
 1. 668
 506
 /note="cap site"
 671..688
 /note="gly4"
 1115..2686
 /gene="gag"
 1115..2686
 /gene="gag"
 /function="retroviral capsid"
 /product="Gag protein"
 join(1115..2686,2690..6271,6150..8132)
 /note="viral polypeptide"
 /codon_start=1
 /protein_id="CAB65340.1"
 /db_xref="GI:6688948"
 /translation="MGQTVTPPLSLTLDHWEYRSRAHNSVQKRGPMOTRCASEMP
 KPRKPGRIILALGKNSAEKVEPSRIPELEPPWPEQVPPVPAQAVRG
 ADIWNKTNHPPSEDFQRLTGIVESIMFSHQPWDCQQLQTLQTLTEERERILLEA
 RKNVPGADGRPTOLNEIDMGFPLTRGMDVNTAEGRRESLKIYRQALVAGLSARRP

TNLAKREVMOGPNPPEVLERLMEAFRRTPPTSEAKASVALATIGSALDIR
 KLIQREGLQENELDLVREAEKYVYRRETEEEKREERERERERDRROBKNIIT
 KILAAVEKSSREERERDRFIRSGPROSGNIGRPTLDKQCVCKEKMAMNCK
 KGNPKVLEBEDKDRGSDPLEPSPVTLKVGQVPELVLDGASHVLLQPLGL
 KEKSGMVGATGQROYPMWTRTRVDGLGRTVSEFLVPECPVLLDRDLTKAGAI
 SEQGRPEVSVNNRILYVITLQIDDEVLYSPOVKPODDISWLMJOPPOMAETAGI
 LAKOVPOVITLQASATVSVKQYTLSEAREGIMPHYORILQOGLIIVPOSPMPL
 LPVRKGTNDYRVDOLREVNKRVODHPVTPNPNLSALPEBRNNTYVLDLDAF
 CLRLHPTSQLPEAFEMDPGRTGQLMTRLPQGFNSPTIFEDALHRLDANKIOH
 POVLLOYVLDLLAGATKODLEGTALLLELSDGLRASAKAQICRRVYVLSGL
 LRGGOWMLTEARKTIVYOIPAPTTAKOYREFLAPGLAALAPLPLETE
 KGEFSAPEHOKAPDAIKKALLSAPALAPVTKPTLYVERGVAGVLTOLGWM
 KRPAVILSKKIDPVASGMPVCLIAAIAVILKADKTLTIGONTIVIAHALENTVO
 PDRWMTNARMHYOSLITERVTPAPRALNPALPLEBEDEVTHEDTLLEEG
 VRKDLIDPLEGEVLTWFTDSSVYVEKRNAGAVNDGPTIMASSIPBETSQKE
 LMAITDIALGEGKSNITDSSVAFATAHAGALYKORGLITSAGRIKKEILSL
 LEALHPKRLAIHICGHOKADLISRONADRYAKOAOAVNLPIIEPKAPER
 ROYLTEDMOETIKKIDOFSETPGCTYSKELPHKGLEVOOIHRLTLGTRKHQ
 OLVRTSPYVHLRPGVADSVYHCYPOQLVNPNRIPKRIKSGHGAHEDFTE
 VPAKYGKNTLLAFVDTISGWAEATPKKSTVYAKKILIEIPRIPRIPVIGSDNG
 PAVAVOSQGLAKILGIMKLHCAIRPOSSGOVERMNTIKETTLKLTETGINDMA
 LLEFVLEFRANTPGQGLNPELYELGSGPPELAVSHASAVVLSQPLESLKLEWY
 ROMAKOLREAYSGEDLOVPHRFQVGSVYRRHAGNLETMRGPVLYLTPTAV
 KYERIPITWISHASHVAPPDSSGMAKENTENPLKLRHVVPYVNNSSMHPILSR
 HLPBGRKPKRLKPLSPASIAMPLITSTPOVNGKRLVDSNSHKPLSLTMLTDSG
 TGINISTOGEAPLGTWPELYVCLRSVTPGLMDATPPDVLARAGVYVCGPPNNE
 YCGNPODFECKQMSCLTSDNGMMKPFVSQODVSTFVNNPTSYOFTYHGRKMDQ
 OYOKDVFNKOISCHSLDLYLKIISTEKKOENLOKWNGLISWGLVYGGSGRKG
 VLFIRLRIETOMPEPAVIGPNKGLAEOGPPIOEOPSPSDNYSSTVSGEKKGS
 KTGAKLESLIOGAFALNSTPEATSSCLASAPYEAGARNGKVNKREHDPII
 TMSGONKLTLEVSQKCTIGMPPSHVCHNTEFNSTSEYOVLVGYRMMACN
 GLTPKSTLYFNQTKPCVMOVIVPPOVHYVPPKVALDYVYVNNKPREPISLTAVM
 LGISVAGVGTGTAALITGPOOLLEKLSMURHIVIEDOALEKSYNLEESLISEV
 VLONRGLDLFLLEKGLCVALKKECCFVVDHSGAIRDSMSLRRLRRRRERERADO
 GMFEGMFRNSPMWTLTSLALGVLVLLLTGVLCLINREVAFVERVSQVJWVLR
 OOOYGLISGETDL"

mat_peptide
 /gene="pol"
 /function="retroviral protease and polymerase"
 /product="Pol protein"
 /gene="pol"
 2690..6271
 2690..6271

gene
 /gene="pol"
 2690..6271
 6150..8129
 /gene="env"
 /function="retroviral envelope"
 /product="Env protein"
 6150..8129
 /gene="env"
 8182..8849
 8742..8747

LTR
 polyA_signal
 8742..8747

BASE COUNT 2355 a 2191 c 2281 g 2022 t

ORIGIN

Query Match 56.5%; Score 4640; DB 14; Length 8849;
 Best local similarity 99.8%; Pred. No. 0;
 Matches 5330; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

Db 1 gtaggtactgactgtagcccaagcgagctggaataaaatcctctgctgttgcac 60
 528 GTGGGTGTACGACTGTGGGCCCGACGCGCTTGGAATAAATCTCTTGCTTTGATC 587
 61 aagacgctctcgtagatgattgggtgtgcgctcttcgagcccgagcagaggggat 120
 588 AAGACCGCTTCTTGATGAGATTGGGGGTGTGCGCTTCCGAGCCGAGAGGGGAT 647
 121 tgtcttactgagccttcaattgtgtcgctggccgggaaatcctgagaccaccccta 180
 648 TGTCTTTACTGCGCTTTCATTTGTGTGTTGGCGGGAATCTGAGCACACCCCTTA 707
 181 caccgagagacgacttgaggtaaagagatccctctggaacgtgtctgtgtcgcgc 240
 708 CACCCGAGAACCGACTTGAGAGTAAAGGATCCCTTTGGAAAGTGTGTGTGTGCGCC 767

241 ggcgtctctgcttctgagtgctgttcttcggtgatcgcgcttccggttgcaagcttcc 300
 768 GGGCTCTCTTCTTGAGTGTGTCTGTTTGGGTATGAGCGGCTTTCGGTTGACGTGCT 827
 301 ctgagacgtaagagacttgagagactgtatcagcaagctgtctaaagagatcaagctg 360
 828 CTCAGACCGTAAGAGACTGAGGAGACTGTGATCAGACACTGCTAGAGAGATCACAGGCT 887
 361 ccaacctggggagcgccccgggaggtgggagagccaggaagcgctggtgtctctact 420
 888 CCACCCCTGGGGACGCCCGGAGGTGGGAGAGCCAGGAGCGCTGAGTGTCTACT 947
 421 gtcgctcagagagacgagctctgttctgaagcgaaagcttccctccggcgctccg 480
 948 GTGCGTGAAGAGACCGAGTTCTGTGTGTAAGCGAAGCTTCCCTCCGCGCGGCTCG 1007
 481 actcttgcctgcttctggaagaagcgagcggtcggtgtgtgtgagctgttggtt 540
 1008 ACTCTTTGCTGCTGTGTGGAAGACGCGGAGCGGTGCGGTGTGTGTGATCTGTGTTT 1067
 541 ctgtctcgtgtcttctgtcttctgtcgtctcttctacagtttaatatggaagacag 600
 1068 CTGTTGTGTGTCTTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1127
 601 tgaactacccctctagtttgaactctgacatctggaactgaagttagaatccaggtcata 660
 1128 TGACTACCCCTCTAGTTTGACTCTGACATCTGACATGTAAGTTAGATCAGGCTCATA 1187
 661 attgtcagctcaggttaagaagagacttggaagacttctgtcctcgaatgagcga 720
 1188 ATTGTCTGCTCAGGTTAAGAGGAGACTTGGAGACTTCTGTGCTCTGTAATGGCCAA 1247
 721 catcgatgttgaatggacatcaagaagagacacttaattctgaatcatctgtctgt 780
 1248 CATTCGATGTTGATGGGCATCAAGGGGAACTTTAATCTGAATTAATCTGCTGTGA 1307
 781 aggaatcatcttcaagactgagacggctctcatctctgctcagggccctatactcta 840
 1308 AGGCATCATTTTTCAGACTGAGCCCGGCTCTCATCTGTGATCAGGAGCCCTTAATCTTA 1367
 841 cgtgcaagatttgcgaagaagatcctccgcatctggttaacaatgactaataaaccga 900
 1368 CGTGCAAGATTTGGCAGAAATCTCTCCGCATGAGTTAAACATGCTAATAATAACCA 1427
 901 gaaagccaggtcccgaaatcctgtgtcttggagagaaacaacaactcgtgcgaagaa 960
 1428 GAAAGCCAGGTCCCGAATCTGCTTGGAGAGAAACAACAACCTCGGCGGAAAAAG 1487
 961 tgaagcctcttctcgtatctaccccgagatcgaggaagccgagacttgccggaacc 1020
 1488 TCGAGCCCTC-TCCTGATCTTACCCCGAGATCAGAGAGCCGCGCACTTGGCCGAAACC 1546
 1021 caactgttcccccaccccttalcacagacagaggtgctgtgagga-cctctgcccct 1079
 1547 CAACCTGTTCCCCACCCCTTATCCAGACAGAGGTGCTGAGAGGACCTTGCCCT 1606
 1080 ccttgaagctcgggtgtgtgagaggaacctctgcggagactcggaacggagagagcgacc 1139
 1607 CCTGAGGCTCCGCTGTGTGAGAGGACTCTCCCGGACTCGGACCGGAGAGGCGCAACC 1666
 1140 ccgagcgagacagagagatcgatctacacgtcgcacactatgagccctcccatgcca 1199
 1667 CCGAGCGGACAGACAGATCGCATTTACCGCTGGCGACACTTATGGCCCTCCCATGCA 1726
 1200 gggggcgaattgcaagccctccagatctgaccttctcttcgaagatctataatttg 1259
 1727 GGGGGCAATTGCAAGCCCTCCAGATATGGCCCTTTTCTTGAGAGATCTTATAAATTGG 1786
 1260 aaactcaacatccccccttctcgaagatccccaagccccaaggggtgtgtgagctc 1319
 1787 AAACCTAACCATCCCTTCTTGAGAGATCCCAACGCTCAGCGGCTTGTGTGAGATCC 1846

OY 1320 ctatgtctctcacagctacttgatgtatgtaacagctgctcacaagctctc 1379
|||||
Db 1847 ctatgtctctcacagctacttgatgtatgtaacagctgctcacaagctctc 1906
|||||
OY 1380 acaacccgagagcagagagaaatctgttaagagctagaanaaatgcttcggggccgac 1439
|||||
Db 1907 ACAACCGAGAGAGAGAGATGTTGTTAGAGGCTAAGAAAAGTTCTCGGGCCGAC 1966
|||||
OY 1440 ggggacccacgcaatgtcaaaatgagattgacatgggaattccctgaactgcgcggt 1499
|||||
Db 1967 GGGCCACCCACCGAGTTCCAAATGATGACATGGATTTCCCTTACTCGCCCGGT 2026
|||||
OY 1500 tggagactacaacagcctgaaaggttagggagcttgaanaactctgcgcagctctgtg 1559
|||||
Db 2027 TGGGACTACAAACAGCGCTGAAGGTAGAGAGCTTGAANAATCTATCGCCAGGCTCTGTG 2086
|||||
OY 1560 gggggtctccggggtcctcaagaagcccaataattgtgtcaaggttaagaaggtgatg 1619
|||||
Db 2087 GCGGGTCTCCGGGGGCTCTCAAGACGGCCACTAATTGGCTAAGGTAAAGAGGTGATG 2146
|||||
OY 1620 caaggacccgaacgaacctccctggtattcttgaagagctcatggaagcttcaagcgg 1679
|||||
Db 2147 CAGGAGCCGAGACGAACCTCCCTCGGTATTTCTTGAAGGCTCATGGAAGCCTTCAGCGCG 2206
|||||
OY 1680 ttaaccccttga tcttaacctcaagagcccaagaagcctcaagtggcctggcctcat 1739
|||||
Db 2207 TTCAACCCCTTTGATCTACTACTAGAGCCCAAGAACCTCAGTGGCCCTTGCCCTTAT 2266
|||||
OY 1740 ggggaagtgcgctctggatatacagaagaacatcaagactggaaaggtttacagaagct 1799
|||||
Db 2267 GGGGAGTGGCTGTGATATCAGAAAGAACTTACAGAGTGAAGGGTTACAGAGAGGT 2226
|||||
OY 1800 gaagtaagctatctagtagagagagcagagaaggtgatatcaagaagggagagaagag 1859
|||||
Db 2327 GAGTTACGTATCTAGTGAAGAGGCGAGAGAAAGGTGATTAACAGAGGGGAGACAGAG 2386
|||||
OY 1860 gagaaggaacagagaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1919
|||||
Db 2387 GAGAGAGACAGAGAAAGAAAGAGAGAGAGAAAGAGAGAAAGAGAGAGAGAGAGAG 2446
|||||
OY 1920 caagagaagaatttgaatgaatcttgccgcagtggttgaaggaagaagcagcagggag 1979
|||||
Db 2447 CAAGAGAGAAATTTGATTAAGATCTTGGCGCGAGTGTGAAGGAGAGAGAGAGAG 2506
|||||
OY 1980 agagagagagattttagaanaatttagtcaagccctagaacatcagggaaactgggcaat 2039
|||||
Db 2507 AGAGAGAGAGATTTTAGAAATTTAGGTAGGCGCTTAGACAGTCAAGGAACTGGGCAAT 2566
|||||
OY 2040 aggaacccactcgaaagaagcagtggtcgtatgttaagaanaaggaacactgggcaag 2099
|||||
Db 2567 AGGACCCCACTCGACAAAGAGACAGTGTGCTATTGTAAGAAAGAAAGAGAGAGAG 2626
|||||
OY 2100 aactgccccaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2159
|||||
Db 2627 AACTGCCCCAAAGAGGAGAAACAAAGAGACGGAAGTCTTAGCTTAGAAGAAATTAAGAT 2686
|||||
OY 2160 taggggagacggtgttcgagacccctcccgagccagaggttaacttgaaggttagagg 2219
|||||
Db 2687 TAGGGAGACGGGTTCCGAGCCCTCCCGAGCCCAAGGTAACTTTTAAAGTTGAGGG 2746
|||||
OY 2220 caacagttgagattccctgtgtgatacaggaagcaggaactcaatgtctgtaacaacta 2279
|||||
Db 2747 CAACAGTTGAGTTCTGTTGATACCGAGAGGAGCATTCAGTCTCTCAACCACTTA 2806
|||||
OY 2280 gaaanaactaaagaanaaanaactcctgggtgtatgggtgcaagaagcaggaatlatca 2339
|||||
Db 2807 GGAANAACATAAAGAAATAATCTGGGTGATGGGTGCAAGAGGCAACGGCAGATATCA 2866
|||||
OY 2340 tggactaccgaagaagcgttgaacttggagtggaagcgttgaacccactgtttctgtgc 2399
|||||
Db 2867 TGGACTACCGAAGAACGTTGACTTGGGAGTGGAGCGGTAACCCACTCTTTCTGTGTC 2926
|||||
OY 2400 atccctgagtgcccaagtaacctcttagtagagaacttaactgaacaaagatlgagactcaa 2459
|||||

Db 2927 ATCCCTGAGTGGCCAGTACCCTTCTGATGAGACTTACTGACCAAGATGGAGCTCA 2986
|||||
OY 2460 atttctttaaacaagaagaaccagaagtgctgtgtaataacaacacactgaactgtgtg 2519
|||||
Db 2987 ATTTCTTTTAAACAAAGAAACCAAGAGTGTGTGTAATTAACCAACCATCACTGTGTG 3046
|||||
OY 2520 accctcaattagatgtatgtatgtacatlatcttccccaaagtaaaagctgaactaagat 2579
|||||
Db 3047 ACCCTCAATTAAGATGATGATATTCGACTATATTCTCCCAAGCTGAATGAAGAT 3106
|||||
OY 2580 ataaagtcctgtgttggagcaggtttccccaagccttgggcaagaacgcaaggaatgggttg 2639
|||||
Db 3107 ATACAGTCTGTTGGAGCAGTTTCCCAAGCCTGGGCGAAGAAACCCAGAGATGGTTTG 3166
|||||
OY 2640 gcaagaagatltcccccaaggttatcaactgaagcagatgtacacagatlatcaagtc 2699
|||||
Db 3167 GCAAGCAAGTTTCCCAAGGTTATTCAACTGAAGGCCAGTCTACACCATATCAGTC 3226
|||||
OY 2700 agacagtaacctgtgttagagagagcttgagaagaagatttggccgagtttcaaaagata 2759
|||||
Db 3227 AGACATACCCCTTGAGTGAAGAGGCTCGAGAAAGAAATTTGGCCCATGTTCAAGATTA 3286
|||||
OY 2760 atccaaggggaalctcctagttccctgtccaatcccttgaatactccctgtacaggtt 2819
|||||
Db 3287 ATCCAAAGAGGATCTCTAGTCTCTGTCATCCCTTGAATATCCCTGCTACCGGT 3346
|||||
OY 2820 aggaagccttgggaaccaatgatatatacgaccaggtacaggaacttgaagaaggtcaataaag 2879
|||||
Db 3347 AGGAAGCTTGGAGCAATATATATGACCAAGTACAGAGACTTGAAGAGGTCAATTAAG 3406
|||||
OY 2880 gtgcaggaatacaaccaaggttcccgaaaccttaaacctcttgaagccctccgct 2939
|||||
Db 3407 GTGCAGAGCATATACCAAGAGTCCCGAAACCTTTAACTCTTGAAGGCTCCCGCT 3466
|||||
OY 2940 gaaaggaactgtgtacagatlatggaacttaagaatgtccttcttgcctgataataac 2999
|||||
Db 3467 GAAGGAATGTGTACAGATTTGACTTAAGATGCTTCTGCTGAGATTAACAC 3526
|||||
OY 3000 cccactagaaccaacacttcttgccttgaatgaggaagatccaagttagggagaagaacggg 3059
|||||
Db 3527 CCGACTAGCCAAACCTTTTGGCTTGAATGAGAGATTCAGAGTACGGGAAGAACCGGG 3586
|||||
OY 3060 cagctacacttgaacccagctgcccgaaggttcaagaactcccgacacatctttagagaa 3119
|||||
Db 3587 CAGCTCACCTGAGACCCGAGCTCCCAAGGTTCAAGAACTCCCGACATCTTTGAGGAA 3646
|||||
OY 3120 gcccatacagaaggaactgtgccaactcaagatccaacacccctcagtgacccctccag 3179
|||||
Db 3647 GCCCTACCAAGGAGCTGGCCCACTTCAAGATCCCAACACCTCAGTGAAGCTCTCCAG 3706
|||||
OY 3180 taagtgaatgaactgtcttgcgagagcacaaccaagaagatcttaagaagtagaag 3239
|||||
Db 3707 TAGGTGATATGACTGCTTGTGCGGAGCCCAACCAAGAGACTGCTTAGAAGGTACCAAG 3766
|||||
OY 3240 gcaactactgtgaaattgtctgaactaggtctacagagcctctgtaagaagggccagatt 3299
|||||
Db 3767 GCATCTACTGTGTGAATGTGCTGACCTTAGCTACAGAGACCTCTGTAAGAAAGGCCAGATT 3826
|||||
OY 3300 tgcagagaagaagtaaaataacttgggtgtacagtttgcggggcgagcagagtgctgacg 3359
|||||
Db 3827 TGCAGGAGAGAGGTAACTATTCTGGGTACAGTTTGGGGGCGGAGGAGATGGCTGACG 3886
|||||
OY 3360 gaggcaaggaagaanaacttgaagcagatccagggcccaacacagcagaagaagagaa 3419
|||||
Db 3887 GAGGCAAGGAAGAAACTGTATGATCCAGATACCGGCCCAACCAAGCAAGAGTGA 3946
|||||
OY 3420 gaatttlttgggaacagctggaatttgcagactgtgataccgggttctgcgaacttagca 3479
|||||
Db 3947 GAGTTTGTGGGAGACCTGTGATTTGACAGCTGTGATCCCGGCTTTGCAACTTAGCA 4006
|||||
OY 3480 gcccaactacacggctaaacaaagaagaaggggaattcctgtggtctcgaagcag 3539
|||||

Db 4007 GCCCACTACCGCTAACCAAGAAAAAGGGAAATCTCTGGGCTCTGAGCACAG 4066
 Qy 3540 aagcatttgatgctatcaaaaagccctgtagcgacccctgctcctgcccctgac 3599
 Db 4067 AAGGATTTGATGCTATCAAAAAGGCCCTGAGGAGACCTGCTCTGGCCCTCCCTGAC 4126
 Qy 3600 gtaactaaaccttaacctatgtgtagtgagcgtlaagggagtagccccagagattta 3659
 Db 4127 GTAACTAAACCCCTTATGCTATGATGAGCGCTAAGGAGTAGGCCGAGAGATTGTA 4186
 Qy 3660 acccaaaccttaggacatggagggagagacctgtgtcctaccctcaagaagcttgaact 3719
 Db 4187 ACCCAACCCCTTATGCTATGATGAGCGCTAAGGAGTAGGCCGAGAGATTGTA 4246
 Qy 3720 gtaagcagtggtgcccctatgtctgtagagctatcgacgctgtagcctactgtgtag 3779
 Db 4247 GTAGCAGTGTGGCCGCTATGTCTGAGGCGCTATGCGACGCTGTGGCCATCTGGTCAAG 4306
 Qy 3780 gacgtgacaaattgactttggagacagaataaactgttaataagcccccatgcatltgag 3839
 Db 4307 GAGCGTACAAATTTGACTTTGGGACAGATATTAAGTAAAGCCCCCATGATTTGGAG 4366
 Qy 3840 aacatgcttcgagcagcccccaacccatgtagtgcacaaagcccgatgacccatcaaa 3899
 Db 4367 AACATGCTTCGCGAGCCCGACAGCGATGATGACCAAGCCCGATGACCCCATATGAA 4426
 Qy 3900 aagctgtcttcacagagaggttcaacttgcctccacagccgctcctcaacctgcaact 3959
 Db 4427 AACCTGCTTCTCAGAGAGAGGCTCACTTTCGCTCCACAGCCGCTCTAACCTCTCCACT 4486
 Qy 3960 cttcgtcctgaaagagatgtagaacagtgactatgtagtccatcaactltgattgag 4019
 Db 4487 CTTTCGCCGGAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4546
 Qy 4020 gtagctgggtgctcgaagagaccttaccagatccgctgtagtggagagtgtagcctgg 4079
 Db 4547 GAGACTGGGGCTCCGAGAGACTTTACAGACATACCGTGTGAGTGGAGAGTCTTAACCTGG 4606
 Qy 4080 ttcactgacggaagcagctatgtgtgtagaggtlaagagatgtagtggagagtgtagt 4139
 Db 4607 TTTACTGACGGAAGCAGCTATGTGTGAGAGGTAAAGAGATGGCTGGGGCGGAGTGTG 4666
 Qy 4140 gacggagcccgacagatctgtagccagcagctgtagcgaagaacttcaagcgcaaaagct 4199
 Db 4667 GAGGGGACCCGACATGTGGGCCGACGCTGGCGGAGAGACTTTCAGCCCAAAAGGCT 4726
 Qy 4200 gagctcagtgccctcagcagcagcttgcgctgtagcgaagaagaatcaataaacttat 4259
 Db 4727 GAGCTCATGGCCCTACCGCAACCTTTGGGGCTGGCCGAGAGGAAATTCATTAACATTTAT 4786
 Qy 4260 acggaacagagtagtgccttgcgactgacacgtagacaggggacatctataaacaag 4319
 Db 4787 ACGGACACAGGATGCTTTGGGACTGCACACGTACACGGGGCATGTATTAACCAAAAG 4846
 Qy 4320 ggtgtgcttactcagcagggaggaataaagaacaagaagaatcttaagcctatta 4379
 Db 4847 GGGTGTCTTACTCAGCAGGAGGAAATTAAGAAACAAAGGAAATTTCTTAAGCCTTTA 4906
 Qy 4380 gaaagcttacttgcgaagaagcttagacttataactgtcctgtagacatcaagaagcc 4439
 Db 4907 GAAAGCTTACATTTGCCAAAAGGCTAGCTTTTATACACTTCTCTGGACATCAAGAAAGCC 4966
 Qy 4440 aagaatcatatctagagggagacagatgtagcaggggttgcgaacagcgagccag 4499
 Db 4967 AAGGATTCATATCTAGAGGAGACCATGCTGACCGGGTGTCCAAAGCAGAGCCAG 5026
 Qy 4500 gctgttaaccttgcctataatagaaagcccaagcccaagagacccagagagctaac 4559
 Db 5027 GCTGTAACCTTCTGCTATATAGAAAGCCCAAGCCCAAGAACCCAGACGACAGTAC 5086
 Qy 4560 accctagaagactgcaagagataaaaaagtagaccagctctctgtagactccgagaggg 4619
 Db 5087 ACCCTAGAGACTGGCAAGATATAAAAGATACCAAGTTCTGTGAGACTCCGAGAGGG 5146

Qy 4620 accgtactacatcatatgggaggaatccctgcccccaagaagggttagaatagtc 4679
 Db 5147 ACCTGCTATACCTATATGGAGGAATCTGCCCCCAAGAAAGAGGTTGAAATATGTC 5206
 Qy 4680 caacagatatacgtlctaacaccactagaagaactaacacccctgacagcttgtagaaga 4739
 Db 5207 CAACGATATACATGCTCACCACCTAGAACCTTAACACCTGCGAGCATTTGTCAGAAAC 5266
 Qy 4740 tccccctatgtctcgaagctaacagagtgagctgtagctcgtgtgtcaaacattgtgtg 4799
 Db 5267 TCCCCATATGCTTCTGAGGCTACAGAGATGGCTGACTGCGTCAAAACATTTGCTG 5326
 Qy 4800 cccctgcaagctggttaagtagcctaccccttcaagaataactccaagaagaagagtaggga 4859
 Db 5327 CCGTCCAGCTGTTAATGCTAATCTTCCAGATACCTCCAGGAAAGAGACTAAGGGGA 5386
 Qy 4860 agccaccagagcgtcactggaagtgagctcactgagtgagtgaaagccgctaaatacga 4919
 Db 5387 AGCCACCCAGCGCTCACTGGAAGTGGAGTCTTCACTAGATTAAGCCGCTAATATAGGA 5446
 Qy 4920 aacaaatactattgtgtttttagaacaacttccagagatgggtagagagcttactact 4979
 Db 5447 AACAAATATCTATTGTTTGTGTGAGACACTTTTCAAGATGGGTAGAGGCTTATCTACT 5506
 Qy 4980 aagaagagacttcaaccgtgtgtgtgaagaataactggaaggaatlttccaagaatt 5039
 Db 5507 AAGAAAGAGACTTAACCGCTGCTGCTAATAAATACTGGAGGAATTTTCCGAGATTT 5566
 Qy 5040 ggaatacctaagtagaagtagtgcagacaatgtagcacttgcgttgcagtagtaagtcag 5099
 Db 5567 GGAATFACCTAAGTAATCGGGTTCAGACATAGTGTCCACTTTCGTGGCCAGGTAAAGTCAG 5526
 Qy 5100 gtagtgcgaagatattggagattgtgtgtagaactgtagtgcattgtgatacagaaccgaagc 5159
 Db 5527 GGACTGGCCAAAGATTTGGGATTTGGAATGGAATGATTTGCTATGATGATGATGATGATG 5686
 Qy 5160 tcaagagagtagaagagtagaagacatlaaagaaccccttaccatattgagcaca 5219
 Db 5687 TCAGAGACAGGTAGAGAGATGATAGAACATTTAAAGAGACCTTACCAATTTGACACCA 5746
 Qy 5220 gtagctgcaatlaagtagtgcctcctcctccttgccttgcctttagtagtagagagac 5279
 Db 5747 GAGACTGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5806
 Qy 5280 acccttggaagatttgggctgagccccctatgaattgcttcaaggggagacccccgttg 5339
 Db 5807 ACCCTGAGACGTTTGGGCTGACCCCTATGAATTTGCTTACCGGGAGACCCCCCGTTG 5866
 Qy 5340 g 5340
 Db 5867 G 5867

RESULT 5
 PEN293657
 LOCUS
 DEFINITION
 and env gene, class B, clone 213.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE

PEN293657 8763 bp DNA VRL 30-MAY-2001
 porcine endogenous type C retrovirus proviral gag gene, pol gene
 and env gene, class B, clone 213.
 AJ293657.1 GI:14275841
 env gene; envelope; gag gene; group specific antigen; pol gene;
 polymerase; polypeptide.
 porcine endogenous type C retrovirus.
 porcine endogenous type C retrovirus.
 Viruses: Retrovirus: Retroviridae; Gammaretrovirus.
 1 (bases 1 to 8763)
 Kirch, U., Fischer, N., Czauderna, F., and Tonjes, R.R.
 Comparison of replication-competent molecular clones of porcine
 endogenous retrovirus class a and class b derived from pig and
 human cells
 J. Virol. 75 (12), 5465-5472 (2001)
 21256017

QY 1153 acgagatcgcatattaccgctgcgaactatgatgcctcccatgcccaggcgcaattgac 1212
|||||
Db 1642 acgagatcgcatattaccgctgcgaactatgatgcctcccatgcccaggcgcaattgac 1701
|||||
QY 1213 agccctccagatattgccccttctctctgcagatctctataattgaaaactaacatc 1272
|||||
Db 1702 accccctccagatattgccccttctctctgcagatctctataattgaaaactaacatc 1761
|||||
QY 1273 ccccttctcgagagatccccaacgctccacggggttggtgagtcctatgtctctc 1332
|||||
Db 1762 ccccttctcgagagatccccaacgctccacggggttggtgagtcctatgtctctc 1821
|||||
QY 1333 accacgctacttggatgatgtacacagctgcgtcagacactctccacacggagagc 1392
|||||
Db 1822 accacgctacttggatgatgtacacagctgcgtcagacactctccacacggagagc 1861
|||||
QY 1393 gagagaagaattctgttaagagctagaaaaaatgtlccctgggacgagcgacccacgc 1452
|||||
Db 1882 gagagaagaattctgttaagagctagaaaaaatgtlccctgggacgagcgacccacgc 1941
|||||
QY 1453 agttgcaaaatgagatgacatggagattccctgacgcccgggttggtgacatacaca 1512
|||||
Db 1942 agttgcaaaatgagatgacatggagattccctgacgcccgggttggtgacatacaca 2001
|||||
QY 1513 cggctgaagtgtaggagagcttgaaaatctatcgccaggtctgttgagcggtctccggg 1572
|||||
Db 2002 cggctgaagtgtaggagagcttgaaaatctatcgccaggtctgttgagcggtctccggg 2061
|||||
QY 1573 ggcgctcaaaagcgcccaactaatttgcttaagtgtaaaagagtgatgcaaggacgaag 1632
|||||
Db 2062 ggcgctcaaaagcgcccaactaatttgcttaagtgtaaaagagtgatgcaaggacgaag 2121
|||||
QY 1633 aacctccctcgatattctcttgagagctcatgtagagccttcaagcggttcaaccctttg 1692
|||||
Db 2122 aacctccctcgatattctcttgagagctcatgtagagccttcaagcggttcaaccctttg 2181
|||||
QY 1693 atctactcagaagccccagaagcctcaatgagctgcgttcaatgtggcagtcggtc 1752
|||||
Db 2182 atctactcagaagccccagaagcctcaatgagctgcgttcaatgtggcagtcggtc 2241
|||||
QY 1753 tggatcacaagaagaacttcaagagcttgaaaggtttacagagagctgagtcggtc 1812
|||||
Db 2242 tggatcacaagaagaacttcaagagcttgaaaggtttacagagagctgagtcggtc 2301
|||||
QY 1813 taatgagaagcgagagaagtggtlatcacagaagggagagacagaagaagaaagacaga 1872
|||||
Db 2302 taatgagaagcgagagaagtggtlatcacagaagggagagacagaagaagaaagacaga 2361
|||||
QY 1873 gaaaagaagaagagagaagaagaagaagaagtgatagacggcaagagaagaatt 1932
|||||
Db 2362 gaaaagaagaagagagaagaagaagaagaagtgatagacggcaagagaagaagaatt 2421
|||||
QY 1933 tgaatagatcttgcgcgagtggttgaagggagagacagcaaggagagagagagatt 1992
|||||
Db 2422 tgaatagatcttgcgcgagtggttgaagggagagacagcaaggagagagagagatt 2481
|||||
QY 1993 ttgagaaaattagtgtaggccttagacagtggaacctgggcaatagaccccatcg 2052
|||||
Db 2482 ttgagaaaattagtgtaggccttagacagtggaacctgggcaatagaccccatcg 2541
|||||
QY 2053 acaagagaccagtggtgctatgttaaaagaaagagacactgggcaaggaactccccaga 2112
|||||
Db 2542 acaagagaccagtggtgctatgttaaaagaaagagacactgggcaaggaactccccaga 2601
|||||
QY 2113 agggaaaacaagagacgaaggtcctagctctagaagaanaataagattagggagagagg 2172
|||||
Db 2602 agggaaaacaagagacgaaggtcctagctctagaagaanaataagattagggagagagg 2661
|||||
QY 2173 gtttcgagacccctcccgagccccaggttaactttgaagtgtaggggcaaccagttgag 2232
|||||
Db 2662 gtttcgagacccctcccgagccccaggttaactttgaagtgtaggggcaaccagttgag 2721
|||||

QY 2233 tccgtgtgatalaccgagcgagagcatcagtgctctacaacocataggaanaataaag 2292
|||||
Db 2722 tccgtgtgtgatalaccgagcgagagcatcagtgctctacaacocataggaanaataaag 2781
|||||
QY 2293 aaaaaaatccctgggtgatggtgtgcacaggggcaagcgagatcgaatgactaacggaa 2352
|||||
Db 2782 aaaaaaatccctgggtgatggtgtgcacaggggcaagcgagatcgaatgactaacggaa 2841
|||||
QY 2353 gaaccgttgacttggagtggaagcggttaacccaactggttcttggtcatcccgatggc 2412
|||||
Db 2842 gaaccgttgacttggagtggaagcggttaacccaactggttcttggtcatcccgatggc 2901
|||||
QY 2413 cagtaaccctcttaagtgtagagacttactgacaaagatggagagctcaaatctcttgaac 2472
|||||
Db 2902 cagtaaccctcttaagtgtagagacttactgacaaagatggagagctcaaatctcttgaac 2961
|||||
QY 2473 aagaaagaccagagatgtctgttaataaacaacccatcactgtgttgagacccatgag 2532
|||||
Db 2962 aagaaagaccagagatgtctgttaataaacaacccatcactgtgttgagacccatgag 3021
|||||
QY 2533 atgataatatacgactatattccccaagtaaaagcttgatcaagaatatacagtcctggt 2592
|||||
Db 3022 atgataatatacgactatattccccaagtaaaagcttgatcaagaatatacagtcctggt 3081
|||||
QY 2593 tggagagaattccccaagcttggtgagcaagaacggcagagatgggttggcaagcaagtc 2652
|||||
Db 3082 tggagagaattccccaagcttggtgagcaagaacggcagagatgggttggcaagcaagtc 3141
|||||
QY 2653 ccccaacaggtlatcactgaagcgagtgctacacagatlatcagtcagagagatccct 2712
|||||
Db 3142 ccccaacaggtlatcactgaagcgagtgctacacagatlatcagtcagagagatccct 3201
|||||
QY 2713 tggatgagaagctcgaagaagaatttggcgcagatgttcaaaagatataccaacaggga 2772
|||||
Db 3202 tggatgagaagctcgaagaagaatttggcgcagatgttcaaaagatataccaacaggga 3261
|||||
QY 2773 tccatgtcctgtcccaatcccttggatatacccttgatccggtttagaagcctggga 2832
|||||
Db 3262 tccatgtcctgtcccaatcccttggatatacccttgatccggtttagaagcctggga 3321
|||||
QY 2833 ccaatgatatacgaccagtagcagagacttgagagagtcataaagaaggtgacagacatac 2892
|||||
Db 3322 ccaatgatatacgaccagtagcagagacttgagagagtcataaagaaggtgacagacatac 3381
|||||
QY 2893 acccaaggttcccgaaaccttataactctttagagcgccctccgcctgaaggaactggt 2952
|||||
Db 3382 acccaaggttcccgaaaccttataactctttagagcgccctccgcctgaaggaactggt 3441
|||||
QY 2953 acacagatattgacttaaaagaatgcctctctcttgcttgatattacaccccatagcaac 3012
|||||
Db 3442 acacagatattgacttaaaagaatgcctctctcttgcttgatattacaccccatagcaac 3501
|||||
QY 3013 caacttlttgctcgaatggagagatccagtgtaggaagaacgggacgtcacacttga 3072
|||||
Db 3502 caacttlttgctcgaatggagagatccagtgtaggaagaacgggacgtcacacttga 3561
|||||
QY 3073 cccgacgtccccaaggttcaagaactcccgagacatcttgaagaaagcctcaacagg 3132
|||||
Db 3562 cccgacgtccccaaggttcaagaactcccgagacatcttgaagaaagcctcaacagg 3621
|||||
QY 3133 acctggcaacttcaagatccacaacccctcaagtgbaacctctcaagtgatgaacc 3192
|||||
Db 3622 acctggcaacttcaagatccacaacccctcaagtgbaacctctcaagtgatgaacc 3681
|||||
QY 3193 tgccttgcggggagccaacaacagagactgtctagaagtgtaggaagcaactactgttg 3252
|||||
Db 3682 tgccttgcggggagccaacaacagagactgtctagaagtgtaggaagcaactactgttg 3741
|||||
QY 3253 aattgtcagactaggtctagaagagctctgttaagaagggcccgagatttgcaagaagag 3312
|||||
Db 3742 aattgtcagactaggtctagaagagctctgttaagaagggcccgagatttgcaagaagag 3801
|||||
QY 3313 taacatacttgggttgaagtttgcgggagcgagcagatgtcgaagagagcaagaga 3372
|||||

|||||
Db 3802 TTAACCTACTGGGTACAGTTTGGCGGGCGGACCGATGGCTGACGGAGGCGACGAGA 3861
OY 3373 aaactgtagtcagataccagcccaacacagcaaaagaatgtagagattttgggga 3432
Db 3862 AAACGTGATGCATACCGGGCCCAACACAGCAAGCAAGTAGAGATTTTGGGGA 3921
OY 3433 cagctgtagtttgcagacttggatccggggttggagacttggagccacttacc 3492
Db 3922 CAGCTGATTTTGGACACTGTGATCCGGGTTTGGACCTTACACGCCACTTAC 3981
OY 3493 ccctaaccaaaaagagggatctctcctggctcctgagcaaccaagaagcatttga 3552
Db 3982 CGCTAACCAAAAGAAAGGGGAATTCCTCGGCTCTGAGACACGAAGGATTTGATG 4041
OY 3553 ctataaaaagccctgctgtagcgcaactgtctgcccctccgtagacttaaccct 3612
Db 4042 CTATCAAAAAGCCCTGTGACCGCACCTGTGCTGCTCCCTGACCTAATAACCT 4101
OY 3613 ttacccttattgtagtagcgttaaggagtagcccgaggattttaaccacacctag 3672
Db 4102 TTACCTTTATGTGATGAGCTTAAGGAGTAGCCCGAGGATTTAACCCAAACCTAG 4161
OY 3673 gaccatgagagagaccctgtgtcctacactgtcaaaaagacttgatctttagccagtg 3732
Db 4162 GACCATGAGAGAGACCTGTGCTGACCTGTCAAAAGAGCTTGATCTGTAGCACTG 4221
OY 3733 ggcctgtatgttgaaggtatcgagctgtgagcattatgttgaagagcttgaacat 3792
Db 4222 GGCCTGATGTGTAAGGCTATGCGAGCTGTGCGCATACTGCTCAAGAGCGCTGCAAT 4281
OY 3793 tgacttggagacagataaactgtlaatalagcccccatgacattggagacatcgttcgc 3852
Db 4282 TGACTTTGGAGAGATATATACCTGTATATAGCCCCCATGTGATGGAAGAACTCGTCCGC 4341
OY 3853 agcccccagacagatgtagatgacaaagcccgatgaccacattaaagccgtcttcca 3912
Db 4342 AGCCCCAGACCGATGATGACCAAGCGCCGATGACCCCATATGAAAGCCGTCTTCTCA 4401
OY 3913 cagaagggctcaacttcgttccacagccgctcctcaacccttccacttcttgcctgaag 3972
Db 4402 CAGAGGGGTCACTTTCGCTCCACAGCGCGCTCAACCCCTGCAACTCTTCCGCTGAAG 4461
OY 3973 agactgtagaacagatgactcatatgttccatcaactatgtttagagagacttgggtcc 4032
Db 4462 AGACTGATGACCAAGTACATGATTTGCTCATCTATTTGATTAAGGAGACTGGGGTCC 4521
OY 4033 gcaagacactiacagatacagcgtgactgtagaagtgctaaactggttcaactgacgaa 4092
Db 4522 GCAGAGACTTACAGACATACCGCTGACTGGAAGTGTCTAACCTGCTTCACTGACGGA 4581
OY 4093 gcaactatgttggaaagtaagaagatggtctggggcgagctgtgtgagagaccgca 4152
Db 4582 GCAGCTATGTGGTAAGGTAAAGATGCTGAGACGGCGGTGGTGGAGCGGACCCGCA 4641
OY 4153 cgaactggccagacagccgctggcggaaggaacttcagcgcaaaagcttggctcagagccc 4212
Db 4642 CGATCTGGGCGACAGCCTGCGCGAAGAACTTCAAGCGGAAAGCTGAGCTCATGGCCC 4701
OY 4213 tcacgcaagccttgcgctgagccgaagggaaatccataaactltatacgacagcaggt 4272
Db 4702 TCACGCAAGCTTGGCGGTGGCGGAGGAAATCCATTAACATTTTATGAGGACAGAGGT 4761
OY 4273 atgcttgcgactgcacagctacaggggccaatataaagaagggttgccttacc 4332
Db 4762 ATGCTTTTGGACTGACACAGTACACGGGGCCATTAATAACAAAGGGGTTGCTTAAC 4821
OY 4333 cagcaggaaggaataaagaacaaaggaatattcaagcctattagaagccttaact 4392
Db 4822 CAGCAGGAGGAAATTAAGAAACAAAGAGAAATTTTAAGCCATTAAGAGCCTTACAT 4881
OY 4393 tgcacaaaagctagctatatataactgtcttgcacatcagaagcacaagatctcat 4452
|||||

Db 4882 TGGCAAAAAGCTAGCTATTATACACTGTCTGACATCGAAGAAAGCCAAAGATCTATAT 4941
OY 4453 ctgagaggaaccagatgagctgacccggttgcgaagcagcagccagcgttataccctc 4512
Db 4942 CTGAGGGGAACCGATGTGCTGACCGGTTTCCAAAGAGGAGCCGACGCTTTAACCTTC 5001
OY 4513 tgcctataatagaacgccccaaagccccagaaaccagacagatcaaccctagaagact 4572
Db 5002 TGCCTATATATAGAAGACCCCAAGCCCAAGACCCAGACAGTACACCTTGAAGACT 5061
OY 4573 ggcgaagagataaaaaagatagaccggttctctgaagctccggagggttgcctataact 4632
Db 5062 GCGAAGAGATATAAATAATGACCATTTCTAGACTCCGGAGGGGACCTGCTATACCT 5121
OY 4633 cataagggaagaatccctccccaagaaggttagaatagttccaaacagatatac 4692
Db 5122 CATATGGGAAGAAATCTCTCCCAAGAAAGGTTAGAAATGTGCCAAGATACAT 5181
OY 4693 gtctaacccacttaggaactaaacactgcagcagttgttcagaacatcccttatactg 4752
Db 5182 GTCTCACCCACCTAGGAATAAACACCTGCAGAGTTGGTCAAGAACATCCCTTATCATG 5241
OY 4753 ttctgaagctaacagagtagctgactcgtgtgttcaacaactgtgtccctgcagctgg 4812
Db 5242 TTCTGAGGCTACAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5301
OY 4813 ttaatgttaactccttccagaatacctccagaagaagactaaggaggagccaccagcg 4872
Db 5302 TTATGTCTAATCTTCCAGATACCTCCAGAAAGAAAGACTAAGGGAGCCACCCAGCGC 5361
OY 4873 ctcaactgggaagtgagcttaactgaggttaagccgggttaatacggaaacaaatattat 4932
Db 5362 CTCACTGGGAAGTGTGACTTCACTGAGGTAAGCGGCTTAATACGGAACAAATATCTAT 5421
OY 4933 tggcttctgtagaaccttctcagatcgggtagaggttlaacctactaaagaagaact 4992
Db 5422 TGGTTTTGTAGACACCTTTTCAAGATGGGAGGCTTATCTTAAGAAAGAGACTT 5481
OY 4993 caaccgtgtgtgtcctaagaataacttgaaggaatatttccaagatttggatccctaaag 5052
Db 5482 CAACTGGGTGCTTAATAAATACTGAGGAAATTTTCCGAGTTGGATTAATCTTAAGG 5541
OY 5053 taataggttcaagacatgctcagacttctgttgcacagtaagtcaagagacttggccaaga 5112
Db 5542 TAACTGGGTGCAAGATGTGCTCACTTTTGTGCCAGGTAAGTCAAGGACTGGCCAGA 5601
OY 5113 tatgggatttgaattggaactgcatgtgtcatcagaaccccaagctcaggaagtag 5172
Db 5602 TATTGGGATGATTGGAACTGCTATGATACAGACCCCAAGCTCAGGACAGGTAG 5661
OY 5173 agagatgaatagacacatlaaagagaccccttaacaaattgacacagagactgca 5332
Db 5662 AGAGATGAATAGAACCATTTAAAGAGACCTTACCAATTTGACCAAGAGACTGGCATTA 5721
OY 5233 atgattgtagtctcctcctccttgccttlttaggttgaaggaacacccctggagact 5292
Db 5722 ATGATTGATAGCTCTCTGCTGCTTGTGTGCTTTTAAAGGTTAGGAACCCCTGAGAGT 5781
OY 5293 ttggcttgaacccctatgaattgtctcagggggaagccccccgcttggagaatctgct 5352
Db 5782 TTGGCTACCCCTGATGAATTTGCTTACGGGGAGACCCCTGCTGTGTAGAAATTTCTT 5841
OY 5353 ttgcacatagctgtagtctgtcttccacagccttgttctcagctcaagcgctcg 5412
Db 5842 CTGTACATATAGTGTGATGTGCTGCTTCCAGCTTTGTCTTGTCTCAAGGCTCAAGGCTCG 5901
OY 5413 agtgggtgagcagcgagctgtgaagcagctccggaggagcttactaagg---aggagact 5469
Db 5902 AGTGGGTAGGCAACGAGCCTGGGAAGCAGCTCCGGAGGCTTACTACGAGAGAGAACT 5961
OY 5470 tgcaggtccacatgccttccaaagttagaagattagctatggttagagcgccacgctgag 5529
Db 5962 TCGAAGTTCCACATCGCTTCAAGTGGAGATTACGCTATGTATGAGACGCCACCATGACG 6021
|||||

QY 5530 gaaacctcgagactcggtggaaggaacctatctgtaactttgacacacacagcgctg 5589
DB 6022 GAACCTCGAGACTCGGTGGAAGGCGCTTATCTGCTACTTTTGNACCAACACAGCGCTG 6081
QY 5590 tgaagtgcgaagaatcccaactgtaacatgacatcccaaglttaagccgacgacgtc 5649
DB 6082 TGAAGTCGAAAGATCTCCACTCGATCATCCACGCTTAAGCTGGCGCCACCTC 6141
QY 5650 cgcagctcgaggtggaagcggaagaagctggaatcccttaagcttgctgcacatcgcc 5709
DB 6142 CCAGCTCGGGGTGGAGAGCGGAAGAGCTGAGATCCCTTAAGCTTCCCTCCATCGCC 6201
QY 5710 tggctccttaacttaacaatactcccaagcgagtagtaaaagccttaagaacgtcg 5769
DB 6202 TGCTTCCTTACTTAACANTAACTCCCGAGCGCAGTAGTAAGAGCGCTTTAGACAGCTCG 6261
QY 5770 aacccccaatagaccttataccttaacctgagctgtaattgacccgtatacgggtgtaact 5829
DB 6262 AACCCCATAGACCTTATCCCTTACCTGGCTGATATTGACCTGATACGGGTGCTCACT 6321
QY 5830 gtaaatagcaactcgaggtgtgtctcttaagaagcgactggtgacgtgaactgacttctgc 5889
DB 6322 GTAATATAGCACTCGAGGTGTCTCTAGAGCACTGTGAGCTGAACTGCAATTCCTGC 6381
QY 5890 ctccgctgataaaccgctgttaagaacacacctcccaactagctcgtatgataag 5949
DB 6382 CTCGCACTGATTAACCCCGCTGTTAAAGCACACCTCCCACTAGTCCGCTGATTAAGG 6441
QY 5950 ttctatgtctgcccagcgacagaagaagaatactgtgaggttctgagggaatctctc 6009
DB 6442 TTCTATTGCTGCCAGGCGACAGAGAAAGAAATACTGTGGGGGTTCTGGGGAATCCTTC 6501
QY 6010 tgaagaagatgaggtcgctgcgtcaacctccaacgataggaactggaatgcccagatctctc 6069
DB 6502 TGTAGAGATGAGGTGCGCTGCTCACTCAACGATGAGAGTGAAGAGCGCATCTCTCTC 6561
QY 6070 caagacgggtaaatcttcccttctgcaatctccgagcgagcgacagataaagataagaa 6129
DB 6562 CAGGACCGGGTAAATTCCTTCTTTGCAATTCGCGCCCGGCAAGTACAAAGTATGATGAA 6621
QY 6130 ctatataagaataagagctgctcccatcaagacttagatctataagaataagtttact 6189
DB 6622 CTATATAAGATTAAGAGCTGCTGCCATCGACTTATGATTAATTAAGATAGTTTCACT 6681
QY 6190 g--aaaggaagaacagaataatctcaaaagtggataaagtgaagctggggaatagt 6247
DB 6682 GAAAAAGGAAACAGGAAATATTTCAAAAGTGATTAATGATGAGCTGGGGAAATAGTT 6741
QY 6248 ttttat-tatatgaggggagagaggttcacacttaacattcogctttagatagagag 6306
DB 6742 TTTTATAATATGCGGGGAGAGGCTCACTTTTAACTTTCCTTGAATGAGAGAG 6801
QY 6307 ggaacagaacccccctgtggaatctggaacccgataaagtaactgcttaacaagggccccg 6366
DB 6802 GGGACAGAAACCCCTGTGGAGTGGGAGCCGATTAAGTACTGGCTTAACAGGGGCCCGG 6861
QY 6367 gccctggagccacgacataacttgcggtgccccaaatlaaactcgtcgtcgccctgaata 6426
DB 6862 GCCCTGAGAGCACGATTAATCTTGGCGTCCCAATTAACCTCGCTGGCGCTGCATA 6921
QY 6427 aacacagcgcttagcaagactgacatgattgattcctcaacaagcgctgaagaatcc 6486
DB 6922 ACACAGCGCGCTTAGCANCGATCACTGATGATTTCTTAACACAGCGCTTGAAGATCTC 6981
QY 6487 ccaaggtcttcgttlaagaacaggaagactcttcagctcatccaaggaagctttccaa 6546
DB 6982 CCAGGTGTTCTGTTAAGACAGAGACAGACTTCAAGTCTCATCCAGGAGACTTTCCAA 7041
QY 6547 gccataaactcaacgagactgatacttcttctgttctgttctatccatcagag 6606
DB 7042 GCCATCAACTCCACGAGACCTGATGCACTTCTTGTGGCTTGTATCTATCCAGG 7101
QY 6607 cctccttatttgaagggaatgctaaagaagaanaattcaatgtgacccaagaacataga 6666
DB 7102 CTTCTTATTAATGAGGGATGAGCTTAAGAGAGAAATTAATCAATGTGACCAAGAGCATAGA 7161
QY 6667 aatcaatgacatggggttcccgaaataagttacccctcaatgaagttccgggaaggg 6726
DB 7162 AATCAATGTACATGGGGGTCCCGAAATGAAGTTACCTCACTGAAGTTCCGGGAAGGG 7221
QY 6727 aatgcataggaagaagctcccccacccacacacacttgcctatagatcgtgttat 6786
DB 7222 ACATCATAGGAAAGCTCCCCCATCCACCAACACCTTTTGCATAGTACTGTGTTAT 7281
QY 6787 gagcagagcctcagaataatcaglatattagtaactgttlaaacaaggtgtgggcaat 6846
DB 7282 GAGCAGGCGCTCAGAAATCAGATATTAGTACCTGGTTATTAACAGGTGGGCAATGCAAT 7341
QY 6847 actggttaacccccctgtgttccactcagcttctcaacaataccaagaattctgtc 6906
DB 7342 ACTGGGTTAACCCCTGTGTTCCACCTCACTCTTCAACCAATCCAAAGATTTCTGTGC 7401
QY 6907 atggtccaaatcgtcccccaggttactacacatccctggaaggtgttctgtgaatat 6966
DB 7402 ATGGTCCAAATCTGCTCCCGAGTGTACTACCATCTGAGAAAGTGTCTTGATGAATAT 7461
QY 6967 gactacggtataaccgaccaaagaagaacggtatcccttaccctagctgtatgctc 7026
DB 7462 GACTATGCGTAAACCGACCAAAAGAGAACCCGATCTCCCTTACCTAGCTGATATGTC 7521
QY 7027 ggaataggaagcgccgttggcgtlaagaacaggaagcttgccttgaatcaagaacag 7086
DB 7522 GGATTAGGAGCGCGCTGTGGGTAGAGAACAGGAGACGTGCTCCGATCAGAGCACAG 7581
QY 7087 caagtaagaagaagcctgtgtgagctatactcggacatgaagaagatctccggacctta 7146
DB 7582 CAGCTAGAGAAAGACTGTGTGAGTACATCGGCGCATGACAGAGATCTCGAGCCTTA 7641
QY 7147 aagagctgttgaacactgaagaagctccctgaacttctgtcgtgaaggtgttctcaag 7206
DB 7642 GAGGAGTCTGTAGCAACCTAGAAAGATGCTCCGATCTTTGTGTGAAGTGGTCTACAG 7701
QY 7207 aacgggaaggaattagatctgctgttcttaagaagaaggttggttlatgtgaacctlaaa 7266
DB 7702 AACCGAGCGGATTTGATCTGCTGTTCTTAAGAGAGGAGTGGGTTATGCGAGCCTTAAA 7761
QY 7267 gaagaatgtcttctatgtatgatactccaagagcctcaagaactcctataaagaact 7326
DB 7762 GAAGAATGTCTTCTTATGTAGATCACTCAGAGGACATCAAGACTCATGAGCAAGCTT 7821
QY 7327 agaaaaagttagagagcgctcgaaggggaagaagagcgtgaaccaaggggtgttgaagga 7386
DB 7822 AGAGAAAGTTTAGAGAGCGCTCGAAGGGAAGAGAGCTGACCAAGGGTGGTTGAAGGA 7881
QY 7387 tggltcaacaggtctccttgaatgaacacacccctgtcttctgtctgaacgggccccctagta 7446
DB 7882 TGGTTCAACAGGTCTCTTGGATGACCAACCTGCTTGTGCTGACAGGGGCCCTTAGTA 7941
QY 7447 gtcctgctcctgttacttaacagtttgggaccttgccttaataatagtttgccttgc 7506
DB 7942 GTCCGTCCTGTTACTTAACAGTTGGGCTTGTCTTAATTAATAGGTTGTTGCTTGT 8001
QY 7507 agagaagatgaggtgagtcagatccagatcagtgtaactgaagaacagatcaacagcctctg 7566
DB 8002 AGAGAACAGTAGTAGCTGCTCAGATCTGTGATTAAGGACACAGTACCAAGGCGCTTCTG 8061
QY 7567 agccaaggaagaactgacacttagccttccagttcctaagaattagaactatlaaagaac 7626
DB 8062 AGCCAAGAGAAACTGACCTTAGCCTTCCAGTTCTTAAGATTAAGACTATTAAACAAGAC 8121
QY 7627 aagaagtgggaatgaagaagttgaanaatgaactaacctcccaagaaccagaagaatga 7686
DB 8122 AAGAAGTGGGAAATCAAGATGAAGAAATGCAACCTTAACCTCCAGAACTCCAGAGAGTTA 8181
QY 7687 ataaaaagctctaaatgccccgaattacagaccctgtgctgcagtaaatagtagaga 7746

```

|||||
Db 8182 ATAAAGCTCTAATGCCCCGGAATTCACACCTGCTGCTCCACCTAATAGTGA 8241
Oy 7747 aggtcaacttccatctgtccagagcctgctatctccgtgaagtaagaacaggaa 7806
Db 8242 AGGTACACTTGTCTATTCTTCCAGGCTGCTATCTGCTAGTAAGATAACAGGAA 8301
Oy 7807 tgaattgactaatcgctatctgattctgtaaaactgactgagcaccatagaagaattga 7866
Db 8302 TGAGTTGACTAATCGCTATCTGATTCTGTAAACGTGACGACCATAGAGAATTTGA 8361
Oy 7867 ttacattgacagccctagctactct 7896
Db 8362 TTACACATTGACGCCCTAGTACTACT 8391

RESULT 6
LOCUS PEN133818 8750 bp DNA VRL 14-APR-2000
DEFINITION Porcine endogenous retrovirus type C proviral gag, pol and env
genes and LTR (class B, clone 43).
ACCESSION AJ133818
VERSION AJ133818.1 GI:6688949
KEYWORDS env gene; env protein; gag gene; gag protein; pol gene; pol
protein.
SOURCE porcine endogenous retrovirus.
ORGANISM porcine endogenous retrovirus.
REFERENCE 1. (bases 1 to 8750)
Cauderina, F., Fischer, N., Boller, K., Kurth, R. and Tonjes, R.R.
Establishment and characterization of molecular clones of porcine
endogenous retroviruses replicating on human cells
J. Virol. 74 (9), 4028-4038 (2000)
MEDLINE 20139394
REFERENCE 2. (bases 1 to 8750)
Tonjes, R.R.
Direct Submission
Submitted (04-MAY-1999) Toenjes R.R., Medical Biotechnology,
Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, Hessen, D-63225
Langen, GERMANY
FEATURES
source location/Qualifiers
1. .8750
/organism="porcine endogenous retrovirus"
/proviral
/strain="type C"
/db_xref="taxon:61673"
/clone="43"
1. .629
LTR 467
misc_feature /note="cap site"
632. .649
primer_bind /note="g14"
1064. .2635
gene /gene="gag"
1064. .2635
mat_peptide /gene="gag"
/function="retroviral capsid"
/product="Gag protein"
join(1064. .2635,2639. .6223,6099. .8072)
CDS /note="viral polyprotein"
/codon_start=1
/protein_id="CAB65341.1"
/db_xref="GI:6688950"
/translation="MGQVTTPLSLTLDHMTVEVSRANHLVQKKGPWOTCASMP
TFDVGMPSEGTENSEIILAVKALIFQTPSSHDDEPIILTMQDLAEDPPWKKPMLN
KPRKGPRIALCEKKHSAKEYEPPRIYPERIEPTPEBPQVPVPPYPAQGANRG
PSAPCPVVEGPAAGTRSRKATPRTDEIALILPLRTGPPPGQQLPLOTWPSS
ADLYNMKTNHPSEDEPRLTGLVESLMSHPTWDCQQLLTFTTEERETILEA
RKNVPADGRPTQONEIDMGFPLTRPGMDYNTAESESLKLYRQALVAGRSARP
TNLAKREVMQGNPEPVSFLERLMEAFRRFPEDPESKORASVALAFIGSALDIR
KTIQRELTQAEFLRDILVREAEKVYRRTEBEKEDRKREKREKREKREKREKREK
KILAAVEEGSSRERERDRKIRISGRSGNIGNRTPLDKDCAYCKEKGHWARNCPK

```

```

KKNKGKYLAEEDKGRGSDLPPEPVALKVEGQPEFLVDGAEHSVLLQPLGKL
KEKKSVMKATQROY PMTTRTYDLAAGRVHSLVLPCEPVLPLGLDILTKMAQI
SFEQGRPEVSVNNKP ITVTLTDLDEEYLSPOVPEPDODIOSMELOFPAEAQIS
LKQVPOVIOAKASATPSVQYVLSSEARGIIPWPHOILIOG ILVPOSPMPTPL
LPRKRGTDVPRVODIREVNRKODIHPTPEVNYNLSALIPERNNWTVLIDAPF
CIRLHPTSPPIFAFMRBDEGTRTQOLWTRLPQGFKASPTIFEDALRDLANRTOH
POVTLIQIVYDILLAGATRKQDCEGTAKLLLESDLG RASAKKAOICRREYLVGS
LGGGQRMVLDEAKRTKVVOIPAPTRKQVREELGTAFORLMPGFATLAAPLYPTKE
KEFESMAPEHOKTFDAIKKALLSAPALPLVTRKPTLYVDKRGAVAGVLOTGLPQ
RRPVLYLSKRDLDEVASGVMVCALAAVALVKKODKLTLCQNTIVIAPHALENVRQ
PDRMNTAARMTHYOSILLTERVTFAPPAALNPATLIPEDDEPYTHCHOLLIEGQ
VKKDLTDIPLITEVLTWPTDSSVYVEKRNAGAAVNDGHTIMASSLPBSTSAKAE
LMAIQLALAEKKSINITYDSRIYAFATVHAGATYKORGLTSGRILKKEETLSL
LEALHLPKRLAIIHCGHQKADLISRONOMADRYAKAQAVALNPLITETPKAPBR
ROYLTLEDMOEIKKIDFSETEGCTYSYGEIILPHKEGLEVOO IHLTLGLTKLPQ
OLVPTSPYHVLRLPGVADSVYKHCYCOLVNAVNSRVYPRGRKRSRSHGAEVDETE
VKPARKYNNKYLILVFDTFSGWVEAYPTKEKSTPAVAKILIEIPRRCIPVYSDNC
PAFVAQVOSGLAKILIGDPMKLCATRPSSQOVRMNTKETITLTKLTTEGIDWIA
LILPEVLPFRNTPGQGLTPYELLTGPPPLVEIASVHSAVLLOSPLSLKALEWY
RORAMKOLREAYSGEGDLOVPHRQVGSYVYRRRAGNLET RKGPLYVLLTPTAV
KVEGISTWIMASHVVLAPPDGSMRAETENPLKRLHRLVYNNNSPGOMHPTLSW
RHLPTRGGEPRKRLRPLSFASIAFELTLITPOASKRILDSNPHRSLSTMILIDP
DGVAVNSTRGVAPRGTMPELHPLRLINDAVKSTPNNLVSYGFYCCPCTEKKC
GSESEPCRRSCVTSNDGKMPISLDDRVKFSYVSTNPKYVMKLYKPKSSPSD
LDYTLISFEKQENIQWIMGSMGIVFKYGGAGSTILIRLITETGPEPVAG
PDKVLAEQPPALPEPPHNLVPOLTSRLPDITQPPSNCTPLIPNIPRNSPVPK
GORLVLIOGAFOAINSDPDATSSCMCLSSGPYEGMAKEGFVNTKHRQCTW
GSRNKLITLLEVSGKTCIGKAPSSHQILCYSEADYQSEMDYLVGYNRMKAMTGL
TPCVASSVEVNOGDCVMVQIAPRVYVHPEEVLVEYQVRYNRPKREVSLLAVNWGL
LGTAVGVTGTAALITGPOLDEGIGELGELHAMTELRALSSEVSLSEVYL
QNRRLDILFLREGGLCAALKEBCFYVDHSGATIDSRSKIRELERRRRERADQCH
FQGMFNRSFMWITLTLALGPLVLVLLLTLPGLPILNRFVAFVREVSAYOIMVLRO
YQGLLSQGETD"
mat_peptide
/gene="pol"
/function="retroviral protease and polymerase"
/gene="pol"
/product="pol protein"
2639. .6223
gene /gene="pol"
6099. .8069
mat_peptide /gene="env"
6099. .8069
gene /function="retroviral envelope"
/product="Env protein"
6099. .8069
LTR /gene="env"
8122. .8750
polya_signal 8643. .8648
BASE COUNT 2288 a 2214 c 2286 g 1962 t
ORIGIN

```

```

Query Match 55.2%; Score 4532; DB 14; Length 8750;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 7862; Conservative 0; Mismatches 51; Indels 8; Gaps 5;
Oy 276 cgcgcttcggttgctgagctgtcctctcagacgtaagactgagagctgtgacagca 335
Db 752 CGCGCTTTCGGTTTGGAGCTGTCTCTCAGACCTGAAGACGTGAGACTGTGATCAGCA 811
Oy 336 gacgttgtagggagatcaaggctgacacccggggagagcccccggggaggttgaggagcc 395
Db 812 GACGTCTAGGAGATCAAGGCTGACCCCTGGGGAGACCCCGGAGGTGGAGAGC 871
Oy 396 caaggagcgcctggtgctccctactctcgltcaagagagacggtctgtgtgaagcga 455
Db 872 CAGGAGAGCCTGGTGTCTCTACTGTGCTGCTGAGAGACGAGCTGTGTGTAAGCGA 931
Oy 456 aagcttccccctccgcgcgcctcagactcttgctcctgtgtggaagacgcgaggt 515
Db 932 AACCTTCCCCCTCCGGCGCCGTCGACCTCTTTCCTGCTTGGAAGACGGAGCGGT 991
Oy 516 cgcgtgtctgagatctgtgtctcctcgtgtgctctgtgtctgtgtgtgtgtcctcgtc 575

```


QY 2735 aattggccgcatgttcaagaattacccaaggacatccctagttccctgtccaaatcccc 2794
| | | | |
Db 3211 AATTGGCCGCGCATGTTTCAAGATTAATCAACAGGGCATCTAGTCTGTCCAAATCCCC 3270
| | | | |
QY 2795 ttggatactccctgctaccggttagaagccttggagccaatgtatttgaaccagtaca 2854
| | | | |
Db 3271 TTGGATACTCCCTGCTACCGGTTAGGAAGCTGGGACCAATATATCAACAGTACA 3330
| | | | |
QY 2855 ggaattgagaaggtcaataaagggtgcagagacatacaccacaagctcccgaaacctta 2914
| | | | |
Db 3331 GGACTTGAAGAGGTCAATTAAGAGGTGACAGACATACACCAAGGTCCGGAACCTTA 3390
| | | | |
QY 2915 taaccttggagcgccctcccgctgaaagcgaactgtgtacaglatattgaaactaaaga 2974
| | | | |
Db 3391 TAACCTTTGAGCGCCCTCCCGCTGGAAGCACTGTGTACACAGTATGTGACTTAAAGA 3450
| | | | |
QY 2975 tgccttcttgcctgaaattacaccccacacagccaacacatttcttccttcgaattgag 3034
| | | | |
Db 3451 TGCCCTTCTTGCCGTGAGATTACACCCACTAGCCACCGCTTTTGGCTTTCGAAATGAG 3510
| | | | |
QY 3035 agatccaggtacggaagaacccgggacgtacacttgaaacccagctgcccgaaggttcaa 3094
| | | | |
Db 3511 AGATCCAGGTACGGGAAGAACCGGGGACGCTACCTGGACCCGACTGGCCCCAAGGTTCAA 3570
| | | | |
QY 3095 gaactcccgacacatctttgacgaagccctacacagagactggtgccaactcaagatcca 3154
| | | | |
Db 3571 GAATCCCCGACCAATCTTTGACGAAGCCCTACACAGGGACCTGGCAACTTCAGAGTCCA 3630
| | | | |
QY 3155 aacacctgaagtgaacccctccctcagtagtgatgacgtcttcttggcgggaagccacaa 3214
| | | | |
Db 3631 ACACCCCTAGGTGACCCCTCCCTCCAGTACGTGATGATGACTGCTTGTGGCGGAGCCACCAA 3690
| | | | |
QY 3215 acaagactgcttagaaggtacgaagacactgctggaattgtctgaacttagctacag 3274
| | | | |
Db 3691 ACAGGACTGCTTAAGAAAGTACGAAGCCACTACTGTAATGTGCTGACCTAGGCTACAG 3750
| | | | |
QY 3275 agccttgcataaagaagccagatcttgcagagaagagtgataacttgggtgatacagttc 3334
| | | | |
Db 3751 AGCCTCTGCTTAGAAGGCCCAATTTGACGAAGAGAGTAACTTGGGGTACAGATT 3810
| | | | |
QY 3335 gcggggcgaggcagcgtatgctgacggaagcagcgaagaagaactgtatccagaataccgc 3394
| | | | |
Db 3811 GCGGGGCGGGCAGGATGGCTGACGGAGGCGACGGAAGAAACGTGATGCCAGTACCAGC 3870
| | | | |
QY 3395 cccaaccacagccaaagaatgagagaatttttggggacagctgtgatttgcagactgtg 3454
| | | | |
Db 3871 CCCAACCAAGCCAAACAAATGAGAGATTTTGGGGACAGCTGATTTGACAGACTGTC 3930
| | | | |
QY 3455 gatcccgagggttgcgacacttagaccccacttaacccgcttaacaaagaagaagggg 3514
| | | | |
Db 3931 GATCCCGGGGTTTGCACACTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAGGGA 3990
| | | | |
QY 3515 attctcctgggtccttgaagcaccagaagcatttgaatgatalcaaaagaagccctgtcag 3574
| | | | |
Db 3991 ATTCTCTGTGGGCTCTGAGACACAGAAACATTTGATGCTATCAAAAAGCCCTGCTGAG 4050
| | | | |
QY 3575 cgaacctgctgtgacctccctgacgttaactaaaccttaaccttattgtgatatgagcg 3634
| | | | |
Db 4051 CGGACTGCTGTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGATGAGCG 4110
| | | | |
QY 3635 taaggagtagcccgagagatttaacccaacaccttagagacatagagagaagccgtgtgc 3694
| | | | |
Db 4111 TAAGGAGTAGCCCGAGAGATTTAACCAAAACCTTAGACATAGAGAGAGACTGTTGC 4170
| | | | |
QY 3695 ctacctgtcaagaagcctgtatccgttgaagcagtggttggccgtatgtctaaagctat 3754
| | | | |
Db 4171 CTACCTGTCAAGAAGCTGTATCTGTAGCCAGTGTGTTGCCGCTATGCTTAAGGCTAT 4230
| | | | |
QY 3755 cgcagctgtgcccatactgtgtcaagagacgttgaacaaatgtaccttgggacagaatataac 3814
| | | | |
Db 4231 CGCAGCTGTGGCCATATCTGTCAAGAGACGCTGACAAATTTGACTTTGGGACACAAATATAC 4290
| | | | |
QY 3815 tgaatagaccccatgcatltgagagaacatgctgcgacgcccagaccgaltgatgac 3874
| | | | |

Db 4291 TGTATATGCCCCCATCATGTAAGAAACATCGTTCCGACACCCCAAGACCATGATGATAC 4350
| | | | |
QY 3875 caacgcccgcatagcccacatacaagcctgtcttctacagaagaaggttcaacttgcctgc 3934
| | | | |
Db 4351 CAACGCCCGCATGACCCACATATCAAGCCGTCTTCTACAGAGAGGCTGACTTGGCTGC 4410
| | | | |
QY 3935 accagcgctctaacctctccactcttcccttgccttgaagagactgtatgaacagttactca 3994
| | | | |
Db 4411 ACCAGCGCTCTAACCTCTCCACTCTTCTGCTGGAAGAGACTGATGAACCACTGACTCA 4470
| | | | |
QY 3995 tgaatgcatcaactatgtattgaagagactgtgtgcgcaagagaccttacaataacc 4054
| | | | |
Db 4471 TGATTGCAATCACTATGTATTGAGAGACTGGGGTCCGCAAGACCTTACAGACTTACC 4530
| | | | |
QY 4055 gctgactggaagaagtgtcaactgtgttcaactgaaggaagcatalgtgtggaagtaa 4114
| | | | |
Db 4531 GCTGACTGGAAGAGTCTTAACCTGTTCACCTGACGGAACAGCTATGTGTGAAGTAA 4590
| | | | |
QY 4115 gaagatgctgtggcgagctgtgtgagcgggaacccgacagatctgggccaagacgtgtgc 4174
| | | | |
Db 4591 GAGGATGTGCTGGGGCGCGGTGTGACCGGACCCACAGATCTGGCCAGACGCTGCGC 4650
| | | | |
QY 4175 ggaagaacttcagcgcaaaagcctgagctcatgtgcccctcaagcgaagcttgcgctgac 4234
| | | | |
Db 4651 GGAAGAACTTCAGCGCAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTCGGCTGCG 4710
| | | | |
QY 4235 cgaagaagaatccataaacaattatagcgaagcagatgactccttgcagctgcacagt 4294
| | | | |
Db 4711 CGAAGGGAATCCATTAACATTTATACGACAGCAGTATGCTTGGCACGACACGT 4770
| | | | |
QY 4295 acaggggccaactataaacaagaagggtgtgttacccttaagaagaaggaaataaaga 4354
| | | | |
Db 4771 ACACGGGGCATCTATATAAACAAGGGGTTGCTTACTCTACGAGGAGGGAATATAAGA 4830
| | | | |
QY 4355 caaagaagaaatcttaagccttataagagccttacaatttgcacaaaaggctagctatat 4414
| | | | |
Db 4831 CAAAGAGAAATTTCTAAGCTATTAAGACCTTACTATTGGCAAAAAGGCTAGCTATTAT 4890
| | | | |
QY 4415 acactgtcctgacatcagaagaagccaaagatcatatcctagaaggaacagatgtcga 4474
| | | | |
Db 4891 ACAGTGTCTTGACATCAAGAAAGCCAAAGATCTCATCTATGAGGGAACAGATGGCTGA 4950
| | | | |
QY 4475 ccgggttgcgaagcagcgccagcctgttaaccttgccttataatagaagcccaa 4534
| | | | |
Db 4951 CCGGGTGGCAAGCAGGAGGCCAGCGTAACTTGTGCTTATATATGAAGACGCCAA 5010
| | | | |
QY 4535 agccccaagaaccagaagcagtagacacccatagaagaactgcaagaagataaagaataga 4594
| | | | |
Db 5011 ACCCCAGAACCAACAGACAGACAGTACACCTAGAAAGACTGCAAGATPAAAAAAGATAGA 5070
| | | | |
QY 4595 ccagttctctgagactccggaggagacgttgtatatacccatataggaagaatcccgcc 4654
| | | | |
Db 5071 CGAGTTCTCTGAGACTCCGGAGGGGACCTGCTATACCTCATATATGGGAAGAAATCTGCC 5130
| | | | |
QY 4655 ccacaagaaggttagaatatgttcaacagatacatcgtctaaaccaactagaactaa 4714
| | | | |
Db 5131 CCACAAGAAGGGTATGATATGTGCCAACATATCATGTGTCAACCCACTAGAACTAA 5190
| | | | |
QY 4715 aaactgtgaagcagttgttcagaacaatcccttcatatgttctgaagctaccagaagtgyc 4774
| | | | |
Db 5191 ACACCTGACAGAGTTGGTCAGAACATCCCTTATCATGTCTGAGCTACAGAGAGTGGC 5250
| | | | |
QY 4775 tgaactggtgttcaaacatgtgtgacctgccaactgttgaatgtatccctccagaat 4834
| | | | |
Db 5251 TGACTCGGTGTCAAACTTTGTGTGCTGCACTGTTATATCTATATCTTCCCAAGT 5310
| | | | |
QY 4835 aacttcagaaagagactaaagggaagccaccccgagcgtcactgtggaagtggaactcac 4894
| | | | |
Db 5311 ACTTCAGAGAAAGAGACTAAAGGAGACCAACCCAGGGGCTCAGCTGGGAATGTGACTTAC 5370
| | | | |
QY 4895 tgaagtaagccggtctaatacggaaacaatatctatgtgttctgttagaaccttctc 4954
| | | | |

Db 5371 TGAGTAAAGCCGGCTAAATACGAAACAATATCTATGTTTGTGACACCTTTTC 5430
Qy 4955 agagtgtagaggttatactactaagaagagacttaacggtggtgtaagaat 5014
Db 5431 AGGATGGGTAGAGCTTATCTACTAAGAAAGAGACTTCAACCGTGTGCTAAAAAAT 5490
Qy 5015 actggaagaatttttccaagattgnaactaactaagatagatagtcacaaatggtcc 5074
Db 5491 ACTGAGAGAAATTTTCCGAGATTGGAATACCTAAGATATCGGGTCAACAAATGTGTC 5550
Qy 5075 agcttgcgtcccaagtaagtaagagagctgcccagaatacttggtgattgattgaaact 5134
Db 5551 AGCTTTGTTGCCAGTAGTACAGGAGCTGCCAAGATATTGGGATTGATTGAAACT 5610
Qy 5135 gcaattggtacacagaccacaaagcttcagagacggttagagagatgaatagaacctta 5194
Db 5611 GCATTGTGATACAGACCCCAAGCTCAGAGAGTAGAGAGATGAATTAACCAATTAA 5670
Qy 5195 agagacccttaacaaattgacacaaagagacttgcatatgattgtagctcctccg 5254
Db 5671 AGAGACCCCTTACCAATTTGACACAGAGACTGGCATTAAATGATTGATACCTCCTGCC 5730
Qy 5255 cttgtgcttttaaggtgaggaacacccctgagacagttgggtgacccctagaatt 5314
Db 5731 CTTTGCTCTTTTAAAGGTTAGGAACACCCCTGACAGATTGGGCTGACCCCTATGAATT 5790
Qy 5315 gctctaaggggagccccccgttggcaagaattgcttgcacatagtgctgctgct 5374
Db 5791 GCTCTACGGGGAGACCCCGGTGGTAGAAATTTGCTGTGATAGTGTCTGATGTGCT 5850
Qy 5375 gcttcccaagccttgtctccttagctagctcaagcgctcagtggtgtagagcagcggt 5434
Db 5851 GCTTCCACAGCTCTGTCTCTAGGCTCAAGGGCGCTGATGGGTGAGGAACAGACGCTG 5910
Qy 5435 gaagcagctccgggagagcctactcaag -- aggaagactgcaagttccacatcgctcca 5491
Db 5911 GAAGCACCTCCGGAGGCTCTACAGAGAGAGAGAGATTGCAAGTTCCACATGCTTTCCA 5970
Qy 5492 agttgagatctgactatgttagacgacacgctgacaggaacctgaagactcggtgga 5551
Db 5971 AGTGGGAGATTCACTGATATTTAGACCCACCGTGCAGAAACCTCGAGCTGGGTGGAA 6030
Qy 5552 gggactatctcgtacttttgacacacaaacagcgctgtgaaagtgcgaagaatcccc 5611
Db 6031 GGGCCCTTATCTGTACTTTTGAACACACACAGCGGTGTAAGTGAAGATCTCCAC 6090
Qy 5612 ctggtatccatgcatcccaagtttaagccgagccacactccgactcggtgtgtagagcg 5671
Db 6091 CTGATTCATGCAATCCACGTTAAAGTGGCGCACCTCCGACTCGGGGGTGGAGACCGA 6150
Qy 5672 aaagactgagaatcccttaagcttgccctcaactgacctggttctactactaacaata 5731
Db 6151 AAAGACTGAGATCCCTTAAAGCTTCGCTCCATCGCTGTGTTCTTACTCTAACATAA 6210
Qy 5732 ctcccaagccagtagtaaacgcttataagacagctcgaaccccccaatagaacttccc 5791
Db 6211 CTCGCCAGGCGAGTAGTAAGACGCTTATAGACAGCTCGAACCCCATAGACCTTATACC 6270
Qy 5792 ttaactgctgattattagaccctgatacgggtgtaactgtaataagactcgagtggtg 5851
Db 6271 TTACTCTGGCTGATTATTAGACCTGATACGGGTGTCACTGTAATATACACACTCGAGGTTGG 6330
Qy 5852 ctcttagagagacctggtgctgaaactgcatcttgccctcgattgattaaaccccggt 5911
Db 6331 CTCTTAAAGGACCTGTGTGGCTGTAAGTGAATTTCTGCTCCGATTGATTAAACCCGCTG 6390
Qy 5912 ttaaaagcacacactcccaactagtcgtagtatagggttctatltgctgccagagacag 5971
Db 6391 TTAAGACACACCTCCCAACTAGTCCGTAGTATTATGGTCTATGCTGACGAGACAG 6450
Qy 5972 agaaagagaataactgtgggggttctggggaatccctctgtaagaatagagctgctga 6031
Db 6451 AGAAGAGAAATACTGTGGGGTTCGTAGGAATCTTCTGTAGAGATGAGACTGGCTGA 6510

Qy 6032 cctccaagatgagagactggaatggtccgactctctccaagaccggglaaaattcct 6091
Db 6511 CCTCCACAGATGAGAGCTGGAATAGCGATCTCTCCAGGACCGGTAATATCTCCT 6570
Qy 6092 ttgtcaattccgcccgggcaagtaacaatgatbaactatataaagataagagctgct 6151
Db 6571 TTGTCAATTCCGGCCGGGCAAGTACAAAGTATGAACTATATTAAGATAAGACTGCT 6630
Qy 6152 ccccatcaactatgatatctaaagataagtttcactg -- aaagaaacaggaataa 6205
Db 6631 CCCCATACACCTTAACTTATCTAAAGATTAAGTTTCACTGAAAAAGAAACGCAAAATA 6690
Qy 6210 ttaaaaagtgaataatglatggaactgaggaatagtttttat - tataatcgaggag 6268
Db 6691 TTCAAAAGTGAATTAATGTTATGACTGGGATATGTTTTTATTAATATGGCGGGAG 6750
Qy 6269 cagggttccacttcaacatctgccttaggatalagaacgaggagacacccctgtgcaa 6328
Db 6751 CAGGGTCCACTTTAAACATTCGCTTAGGATAGAGACGGGACAGACCCCTGTGGCAG 6810
Qy 6329 tggagccgataaagtagctgtaacaggggccccggccctgtagccacgctaact 6388
Db 6811 TGGGACCCGATTAAGTACTGCTGAAACAGGGGCCCGCCCTGAGACACCGCATTAAT 6870
Qy 6389 tgcggtgcccacattaaactctgctgagccctgacataacacagcgcctagcaacagta 6448
Db 6871 TGCCTGCTCCCAATTAACCTGCTGCGGCTGACATATACAGCCGCTAGCAACGGTA 6930
Qy 6449 ccactgtagtatactccacaaacagcctlagaacatcccaaggtgttctcgtttaagaag 6508
Db 6931 CCACGTGATGATTCTCTACCAACAGCGCTAGAAATCCCGACGTCTCTGTTAAGACAG 6990
Qy 6509 gacagaagacttcaagttcatccaaggagctttccaagccataacatccacagacctg 6568
Db 6991 GACAGAGACTCTTCACTGATCATCAGGAGAGCTTTCCAAACCAATCCACACCGACCTG 7050
Qy 6569 atgcacacttctctgttggcttgcatactcaagagccctcattatagagggatg 6628
Db 7051 ATGCCACTCTCTTGT 7110
Qy 6629 ctaagaagaagaattcaatgtgaccaaagagcatagaatcatatgataatggtggctcc 6688
Db 7111 CTAAGAAAGAAATTAATGATGACCAAAAGACATAGAAATCAATGATGGGGTCC 7170
Qy 6689 gaataatgattccctcaactgaagtttccgggaaggagactgcatagaagaagtcccc 6748
Db 7171 GAAATTAAGCTTACCTCACTGAAGTTTCCGGGAAGGACATGACATAGGAAAGCTCC 7230
Qy 6749 catcccaacaacacttgcataagtagtgcattatgtagagagcctcaagaataact 6808
Db 7231 CATCCACCAACACCTTCTGCTATGATGATGCTGTGTGTGTGTGTGTGTGTGTGTGT 7290
Qy 6809 attagtagctggttataaacaaggtggtggaatgcaatactggtttaaaccctgtgtt 6868
Db 7291 ATTATGATCTGTTATTAACAGGTGTGCGCATGCAATACCTGGTTAAACCCCTGTGTT 7350
Qy 6869 ccacctcagcttcaacaacatccaaagtttctggtatggttccaatctgccccag 6928
Db 7351 CCTCTCACTTCAACCAATCCAAAGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7410
Qy 6929 tgaactacatcctgaagagtagtgcctgtaataatgataatgataatgataatgataat 6988
Db 7411 TGATACATCTCTAGAGAGT 7470
Qy 6989 aaagagaacglatcccttaccctagctgtaatgctcggattaggaacggcgttggcg 7048
Db 7471 AAAGAGAACCCCTATCCCTTACCTAGCTGTATCTGTGTGTGTGTGTGTGTGTGTGT 7530
Qy 7049 taggaacggagacgtgcccgtgatacagaagcccaacagctagaagaagagactgtg 7108
Db 7531 TAGGAACAGGACAGCTGCTGATCAGGACACAGCAGCAGCTAGAGAAAGACTTGGTG 7590


```
WKKFFRDLEYLR"
gene      5294. .7267
          /gene="env"
CDS       5294. .7267
```

```

/codon_start=1
/protein_id="CA76583.1"
/db_xref="GI:3005988"
/db_xref="SPTMBL:O73506"
/translation="MHPTLWRHLPTNGGHPKRLRIPLSFASIAWFLTLITTPQASSK
RLTIDSNPHRPLSLTWTLLIDPDTGVTVNSTRGVAAPGTWMPDLHGLRLINPAVKSP
PMLVSYGFPCCPGTEKEKEKYGCGSEFSCRWGCVISNGDKWPIISLQDVKRSPFN
SPBGYKVMKLIYDKRCSQSPBLDYRLISPEKKOENQILQKINMSGWSIVYKXGGA
GSTLLRLRIETGTEPEPPAAGDPKVLAEQGPALPEPHLIPCOLISLIPITPSPEN
GTTTGLPTNTPRNSGVPVKTGORTLSILOGAQAINTSDPATISCMILCISGPIT
EEMAEKGFNTEKRRKQVCKSGSKNKLITLTVYSKCTIGKAPSPHQHCLSTVYDQ
ASENQYLYPVRNMAACNTGLTPVASSVFNSQKDCVQIVPVPVYHPEEVLVDY
DYRYNRKREPVSLTILAVMLQGTAGVGVGTALITGQOOLEKGLSIHAAAMEDIR
ALAEBSVLSLESLTVLQNRGDLIFLREGGLCALEKECCFVDSHSGAIIRS
MKLELRLERRRRREADQWEGFWNRSWMTTLLSALTGLVTLVLLLTIVGECILN
RVAVARRGAVQIIVLROOYGLLSQCEITD"

```

Query Match	49.98;	Score 4096;	DB 14;	Length 7808;
Best Local Similarity	99.28;	Pred. No. 0;		
Matches 7746; Conservative	0;	Mismatches 55;	Indels 9;	Gaps 6;

QY	329	atcagcaagcgcgtctagagagatcaacagctcgcaccccttggggagagcccggaagttg	388
Db	1	ATCAGAGAGACGTGTAGAGAGATCACAGGCTGCCACCTTGGGGAGCGCCGGGAGGTGG	60
QY	389	ggaagaccagaggagccctgtgtgtctccctactgttcgcagaagaccagttctgtgtc	448
Db	61	GGAGAGCCAGGGAGCCCTGTGGTGTCTCTACTGTGGTCAAGAGACCGAGTTCGTTGTT	120
QY	449	gaagagaaagcttcccccctccgggcgcgtccgactctttgcctgtctgtggaagcgcg	508
Db	121	GAAGCGAAAGCTTCCCCCTCCGCGGCGCTCCGACTCTTTTGGCCGTGTGGAAAGACGC	180
QY	509	gaacggctgcgtgtgtctgcatctgttgccttgcctcgcgtgtctcttctgtgtcgt	568
Db	181	GACGGGTGCTGTGTGTGTGATCTGTTGGTTTCGTTTGGTGTCTTTGTCTTTGTGGCT	240
QY	569	ccctgtctacagttttaatatrgggacagacagtgactaccccccttaagtttgactcga	628
Db	241	CGTTGTCTACAGTTTAAATATATGGACAGACAGTGCACACCCCTTAGTTAACTCGA	300
QY	629	ccattgagactaagtttaagatccaggagctcaaatcttgaagttcagtttaagaaggacc	688
Db	301	CCATTGGACTGAACTTAATATCCAGAGGCTCATTAATTTCTGATTCAGGTTAAGAAGGAC	360
QY	689	ttggcagacttctgtgcctctgtaatgtggcaacatctgcatgttgaaatggcaccagaag	748
Db	361	TTGGCAGACTTCTGTGCTCTGTGAATGGCAACATTGTGATGTTGATGGCATCAGGG	420
QY	749	gaccttaattctiaaattatcccggtgtgtaaggcaatcaattcttcagactgagccgg	808
Db	421	GACCTTAATTCTAATAATATCCGTGGGTGTTAAGGCATATCTTTACAGCTGGACCCAG	480
QY	809	ctctcatctcgtatcagaagccctatccttaacgttgcaagatttggagaagaatcttc	868
Db	481	CTCTCATCTCTGATCAGGAGCCCTAATCTTACTGTGCAAGACTTGGCAGAAATCTCC	540
QY	869	gccatlggttaaccatcgtcgttaataaaccagaagaagccaggtcccgatcctgctct	928
Db	541	GCCATGGGTAAACCATGGCTAAATAATAACCAAGAACCCAGTCCGCCGAATCCTGGCTT	600
QY	929	ttgagagaaaaaanaaacctcgcgcgaaaaaagtcgagccctctctctgtatctaccgcg	988
Db	601	TGGGAGAAAAACAACTACTCTGGCGAAAAAGTCGAGCCCTC-TCTGTGATCTACCCCG	659

QY	989	aalcagagagacgcgcgaatttgcgcggaaaccccaactgttccccaacccctatcaag	1048
Db	660	AGATCGAGGAGCCGCCGACTTGTGGCCGGAACCCCAACCTGTTTCCCAACCCCTTATCCAG	719
QY	1049	cacaggttctgtgagagga - cctctgcacctcttgagctccgggtggtgagagacctg	1107
Db	720	CACAGGATGTGTGAGGGGAACTCTGGCCCTCCTTGSAGCTCCGGTGTGTGAGGAGACTG	779
QY	1108	ctgcgcggaactgcgagccggaagagcgcaaccccggaagcggacagacagatcgcgatat	1167
Db	780	CTCCCGGAGCTCGGAGCCGGAGAGCGGCAACCCCGGAGCGGACAGAGATCGGATAT	839
QY	1168	tacgcctgcgaacctatgacctcccaatgcgcgaaggggacattgagcccccaattt	1227
Db	840	TACCGTGGCAACATGATGCCCCCTCCATGCGGGGGGCCAATTGGCAAGCCCCCTCGATTT	899
QY	1228	ggcccccttctctcgcagatcctatataattggaataactaacatcccccttctcgaag	1287
Db	900	GGCCCCCTTCTCTCTCAGATCTATTAATGGAAACTTAACATCCCCCTTCTCGAGG	959
QY	1288	atcccaagcgtcaacggggttctgtgagcctctatgtctctcaacgactacttgg	1347
Db	960	ATCCCAAGCCTCACGGGGTTGGTGTGAGTCCCTTATGTTCTTCAACAGCCTTACTTGGG	1019
QY	1348	atgatgtcaacagctctgtcagaacctcttcaaacagagagcgagagagaaattcgt	1407
Db	1020	ATGATGTCAACAGCTCGTCGACAGACTCTTCAACAGCCGGAGGAGAGGAATTCGT	1079
QY	1408	taagagcttagaaaaaatttctctgaggcgaagcgagcccaacgcagttgcaaaataga	1467
Db	1080	TAGAGGCTTAGAAAAATGTTCTTGGGGCGACGGCGACCCACGAGATTGGAAAAATGGA	1139
QY	1468	ttagaatggaatttccctctgactgcgcccggtttggagtactacaacacggctgaagttag	1527
Db	1140	TTAGATATGGGATTTCCCTTGACTCGCTGCGGTTGGGACTTACAACAGGCTGAAGTAAAG	1199
QY	1528	agaagcttgaaaatctatctgcgcaggtcttgtgtgcgagctccggggcgtccaagacgc	1587
Db	1200	AGAGCTTAGAAATCTATCGCCAGGAGCTCTGTGGCGGGGTCTCCGGGGCCCTCAAGAGCGC	1259
QY	1588	ccaactaatcttgctaagtagaagagtgtgaagagacggaagaaactctcccgctat	1647
Db	1260	CCACTAATTTGGCTAAGTAAAGAGGTATGAGGGAGCCGACCACTCCCTCGGATAT	1319
QY	1648	ttctctgagaagctcaatggaagcctcaagcgcgttcaaccccttctatcactcaactcaag	1707
Db	1320	TTCTTTGAGAGGCTCATGGAAGCCCTTCAAGGGGCTTCAACCCCTTTGATCTCTACTCGGAG	1379
QY	1708	cccagaagccttaagtgcctctgcctctaatgggcagttcggtcttctgtatatcaagaaga	1767
Db	1380	CCCAGAAAGCCTCATGTGGCCCTTATTTAGGGAGTGGGCTCTGGATTTCTGAAGAAGA	1439
QY	1768	aactcagaagacttgaagaggttlaacagagagctgagttlaagttatcagtagtgaagagcag	1827
Db	1440	AACCTTCAAGAGCTGGAAGGGTCTTACAGAGAGCTGAGTTACGATCTAGTGAAGAGGCA	1499
QY	1828	agaaggttatttaagaagaggagacgaagagagagagagagagagagagaaagaaagaga	1887
Db	1500	AGAAAGCTATTTCGGAGGAGGAGACAGAAAGAGGAGAAAGAACAGAAAGAAAGAGAGA	1559
QY	1888	gagaagaagaggagaaagacggtgatalagacggaagagaaatcttactaaagattcttg	1947
Db	1560	GAGAAAGAAAGGAGGAAAGCGTGAATAGACGCGCAGAGAAATAATTTGACTTAATCTTGG	1619
QY	1948	ccgcaggtgtgttgaagggaagacgacgcagggagagagagagatcttaagaaataagtt	2007
Db	1620	CCGCAAGTGTGTTGAAGGGAAGAGACAGAGGAGAGAGAGATTTTATGAGAAATTAAGST	1679
QY	2008	caggcccttagaacgttcaaggaaccttgggaaatagaacccaactctgaacaaagacagttgt	2067
Db	1680	CAGGCGCTTAACACAGTCAAGGAACCTGGGCAATAGGACCCCACTCGAGAAAGACCAATGTG	1739
QY	2068	cgatctglaaagaaaaagacacttggcaagagaaacttcccagaaggaaggaacaaagagc	2127

Db	1740	CGATTCTGAAGAGAAAGGACACTGCGGCAAGAACTGCCCCAGAGAGGAAACAAAGAC	1799
Qy	2128	cgaaggtccctaagctctagaagaaga taaaga ttagggagacggygttcgacccctcc	2187
Db	1800	CGAAGTCTCTAGCTAGAGAGAAAGATTAAGATTGAGGAGACGGGGTTTCGGACCCCTCC	1859
Qy	2188	cgaagcccgaggttaacttgaagtgaggtgggggacccaagtgtggttcgtttgatacgg	2247
Db	1860	CTAGAGCCAGGGGTAACTTTTGAAGGTGAGGGGCAACCA GTTGA GTTCTTGTTGATACGG	1919
Qy	2248	gagcggagcaattcaagtgctgtcacacaat tagga aaaa ctlaa gaaaaa aaaa tctctgg	2307
Db	1920	GACGGAGCATTCAGTGCCTGCTACACAACATTAGGAAAC TTAAGGAAAAATCCGTGG	1979
Qy	2308	tgatgggtgcacaagggcaagggcagat lcaatgta cta cccgaagaacggttgacttgg	2367
Db	1980	TGTGGGTGGCCACAGGGCAACGGCAGTATCCATGTGAC TACCCGGAAGAACGTTGACTTGG	2039
Qy	2368	gaatggagcaggtgaacccaactggttcttggtatccctggatggcgaataccctcttag	2427
Db	2040	CAGTGGAGACGGGTAAACCCACTGgttttctgtatctccgtagttgccagtaaccttcttag	2099
Qy	2428	gtagaagcttaacgcacaagat tggagatc caaattcttcttgaacaaaggaaacagaag	2487
Db	2100	GTAGAGACTTACTGACCAAGATGGGAGCTCAATTTCTTTGAACAAGGAAGACCAAG	2159
Qy	2488	tgtctgtgaataacaaaccca lcaactggtgttgacctccaattgatatgataatagac	2547
Db	2160	TGCTGTGATTAACAAACCCATCAGTGTGTTAACCTCCCAATGATGATGATATGAC	2219
Qy	2548	tatacttcccgaagtlaaagcctgatacaagatala cag tctctgtt tggagcagttcccc	2607
Db	2220	TATTTCTCCCAAGTAAAGCCGTGATCAAGATATACGTCCTGGTTGGAGCAGTTTCCC	2279
Qy	2608	aagccgtggcagaacacggcagggatggtgttggcaagcaagttcccccaagttatc	2667
Db	2280	AAGCCTGGGAGAAACCGCAGGGATGGTTTGGCAAGCAAGTCCCCACAGGTATATC	2339
Qy	2668	aactggaagccagtgactcaacacagta tcaagtacagaacgtacccctttagtaagaagctc	2727
Db	2340	AACGTAAAGGCCAGTGCCTACACACAGTATCAGTCA GACAGATACCCCTTGAGTAGAAGCTC	2399
Qy	2728	gagaagaatttggcgcagatgttcaaa gattaatcca aagggatcctagtctctgtcc	2787
Db	2400	GAGAGGAAATTTGGCCGCATGTTCAAAAGATTATCAACAGGGCACTCTAGTCTGTGCC	2459
Qy	2788	aatcccttggataatcccccctgtcaacggttaggaagccttggacccaatgatatcgac	2847
Db	2460	AATCCCTTGGAAATCTCCCTGCTACCGGTTAAGAAAGCCTGGGACAAATGATTATGAC	2519
Qy	2848	cagtaacagacttgaagagagttcaataaaagg tlcaggatatacaccacaagctccga	2907
Db	2520	CAGTACAGAGCTTAGAGAGGTCAATTAAGAGGTGACAGACATACACCAAGGGTCCGA	2579
Qy	2908	accctataacctttagagccctcccgctcaacggagacgtgtgaacagatattgact	2967
Db	2580	ACCTTATTAACCTTTGAGGCGCCCTCCGCTGAAGCAACTGTGATACAGATTGGACT	2639
Qy	2968	taaaagatgaccttcttcgctgagattaaacccaactagcaaaccaacttlttgctctcg	3027
Db	2640	TAAAGATGCGCTTCTTGCGCTGAGATTACACCCCACTAGCAACACACTTTTGCTCTCG	2699
Qy	3028	aatgtagagatcccaagta cgggaagacggcagctacactccttggaccccgactgcgccaag	3087
Db	2700	AATGAGAGATCCAGGTACGGGAAGAACCGGGAGCTCACCTTGAGACCCGACTGCCCAAG	2758
Qy	3088	agttcaagaactcccgacacactttagcaagcccttaacacagagacgtgcgaacttca	3147
Db	2760	GGTTAAGAACTCCCGACACTTTTTCACGAAGCCCTACACAGGAGACGTGGCAACTTCA	2819
Qy	3148	ggatccaacacctcaagtgagccctccagta cgtgtgactgtcttcgycggag	3207

Db	2820	GGATCCAAACACCCCTCAGGTGACCCCTCCCTCCACGTACAGTGATGATGACGTGCTTCTGGCGGGAG	2879
Qy	3208	ccacccaacaaggagctgcttagaaggtacgaaggcactactgcttggaaattgcttgtaacctag	3267
Db	2880	CCACCCAAACAGGAGCTGCTTAGAAGGTACGAAGGCGCATCTACGTCTGGAAATTGCTGACCTTAG	2939
Qy	3268	gctcacgagcgtctgcttaagaaggccacagatttcagagagaggttaacatactcttgggtc	3327
Db	2940	GCTACAGAGCTCTGCTTAAGAAGGCCCGAGATTTTCAGAGAGAGAGTACCTACTTGGGGT	2999
Qy	3328	acagatttcgcyggcgcygcagctgagctgacgagagcgacggaagaanaactgtagtcacga	3387
Db	3000	ACAGTTTGGCGGGCGGGCGAGCATGGCTGACGAGGCGACGGAAAGAAAAGCTGTGTCACAA	3059
Qy	3388	taccgcgcccaacacacagccaacaaagtggaagagtttttgggaagacgttgatttcga	3447
Db	3060	TACCGGCCCCAACCCACAGCCCAACAAAGTgagagacttttttgggacagcttgatttttcga	3119
Qy	3448	gactgtgatacccgagggttttgagaccttagacagccccactacacgtctaacccaagaana	3507
Db	3120	GACTGTGATCCCGGGGGTTTGGACCTTAGCGAGCCCCACTCTACCCGCTAACCAAGAA	3179
Qy	3508	aagggggattctccctggtgctctgagcaccgaaagagcatgtgagctataaagaagccc	3567
Db	3180	AAGGGGAAATTCCTCTGGGCTCTGAGCACACAAAGCATTTGATGCTATCAAAAGGCC	3239
Qy	3568	tgctgagcagcaactgtctctgagccctccctcgagctaacaaaccttacccttattgtgg	3627
Db	3240	TGCTGACAGCAGCTGCTGTGGGCTCCCTCGAGCTAACTAAACCTTTACCTTTATGTGG	3299
Qy	3628	atgagcgtaaggagtagcccgagagattttaaaccaaaccttgagccatgtgagagagag	3687
Db	3300	ATGAGCGCTAAGGGAGTAGCCCGAGGGATTTTAACCAAAACCTCTAGGACCAATGGAGGAGAC	3359
Qy	3688	ctgttgcttaacctgctcaagaagctgtgacccgttagcagtggtctggacccgtatgtctga	3747
Db	3360	CTGTGGCTTACTCTGTCAAAAGACTTGAATCGTGAAGCCAGTGTGGGCGGTAGTGGCTGTA	3419
Qy	3748	aggctatcogcagctgtgtgcccatactgtgtcaagaacgctgcgaacaatttgactttggagaga	3807
Db	3420	AGGCTATCAGCAGCTGTGGCATCTGTGGTAAGGAGCGTACAAATGACCTTGGGACAGA	3479
Qy	3808	ataataactgttaataagcccccaatgcatacttggaagaacatcgtctcggagacccccaacgat	3867
Db	3480	ATATTAATCTGTAATAGCCCCCAATGATGTGGAGACATGTTGGCAGCCCCCAACCGAT	3539
Qy	3868	ggaatgaccaacgcccgcgacatagccactatcaaaagctgcttctcaacgaagaggttaact	3927
Db	3540	GGATATCCCAACGCCCGCATACCCACTATCAAAAGCTGCTTCTCAACAAGAGGGTCACTT	3599
Qy	3928	tcggtcccaacgagcgtctcaaacccctgcacactctctgcttgaagaagactatgataacag	3987
Db	3600	TGCTTCACACGCGGCTCTACCCCTTGCCACTCTTCTGCTGAAGACGTGATGATACAG	3659
Qy	3988	tgactactgatttgcacataactatgtattgaaggagactgtggtctccgaagaagacttaccag	4047
Db	3660	TCAGCTCAGATTGGCATCACTATTGATTGAGAGAGCTGGGCTCGCAGAGACCTTTACAG	3719
Qy	4048	acataccgctgacgtgaggaagatgtaacctggtcttaactcgaagaagacgactatgtgtgg	4107
Db	3720	ACATATCCCTCTACGTGGAAAGTGTACTACCTGGTTCCTACACGGAAGCACTATGTGGTGG	3779
Qy	4108	aaggtaaagaagatgtgcctggggcgagtggttgagcgggacccgcagactctggtggcagca	4167
Db	3780	AAGGTAAAGAGGTGGCTGGGGCGGCGGTGTGAGACGGGACCCACAGATCTTGGGCACCA	3839
Qy	4168	gccgtccggaaggaacttcagcgcgaagaagcttagtctcattgaccttaacgaagaagtttgc	4227
Db	3840	GGCTCCGGGAAGGAATTCACGCAAAAGGCTAGCTATGGCCCTTCACGCAAGCTTTTGC	3899
Qy	4228	ggctgcccgaagaagaaatccataaacttataccgagacagcagttatgcttcttgcgactg	4287
Db	3900	GGCTGGCCGAAGGAATTCATTAACATTTATACGGGACAGCAGGTATGCTTTGGCAGCTG	3959

QY 4288 cacaaglacacgagggccatctataaacaagggtgtgtcttaccgaagcaggaggaagaa 4347
DB 3960 CACACGTACACAGGGCCATCTATAAACAAAGGGGGTGTGCTTACCTCAGCAGGGAGGGAAA 4019
QY 4348 taaagaacaagaaggaaattctaagcctattagaagccttacaattgtgccaanaaggtctg 4407
DB 4020 TAAAGAACAAAGAGGAATTTAGAGCTATTAGAGCTTACATTGTCCTCAAAAAAGGCTAG 4079
QY 4408 ctattataactgtctctgacatcaagaagaaccagaatctcatatctagaagggaacaga 4467
DB 4480 CTATTATACACTGTCTCGACATCAGAACCAAGATCTCATATCTAGAGGAGACACAGA 4139
QY 4468 tggctgacgggtgtccaaagcaggccaggctgttacccttctgtgcataatagaagaa 4527
DB 4140 TGGCTGACCGGGTGTCCAAACAGCAGCAGCCAGGCTGTAACTTCCTGCTATATAGAAA 4199
QY 4528 cgcccaagagcccaagaccagacagtlacacccctagaagaactgtgcaagagataaaaa 4587
DB 4200 CGCCCAAGCCCAAGAACCCAGACAGACATACACCTAGAAAGACTGSCAAGAGATAAAA 4259
QY 4588 agatagaaccagttctctgagactccggaggagacctgtctatcctaatactgggaagaa 4647
DB 4260 AGATAGACCAATCTCTGAGACTCCGGAGGGAGCTGTATACCTCATATGAGGAGGAAA 4319
QY 4648 tctgtccccaagaagaagggttagaataatgtccaaacagatacactgtcttaccactag 4707
DB 4320 TCTGTCCCCCAAGAAAGGGGTAGAAATATGTCCAAACAGATACATGTCTCACCCTAG 4379
QY 4708 gaactaaacacctgacagactgtgtcagaacatcccttcaatgtctctagagctacag 4767
DB 4380 GAACTAAACACTGACAGCAGTGTGTCCAGAAATCCCTTATCATGTCTCAGGCTACAG 4439
QY 4768 gagtgcctgactcgtgtgcaacaactgtgtccctgtccagctgtgttaagttaacctc 4827
DB 4440 GAGTGGCTGACTCGTGTGTCAAACTGTGTGTGCTCCCTGCTCAGCTGTTAATGCTAATCCT 4499
QY 4828 ccagaataccctccaggaagaagactaagggaagcccaagggcgcactcgggaagtg 4887
DB 4500 CCAGATATCTCCAGAAAGAGACTTAAGGGAGACCCAGGGGCTCCTCAGGGAATGG 4559
QY 4888 acttaactgagtaaaagccgactaatacagaacaataatclatttgttttcttgaaga 4947
DB 4560 ACTTCACTGAGSTAAAGCCGGCTAATATGSAACAAATATCTATGTTTGTATAGCA 4619
QY 4948 ccttttcagagatggtagaaggctatacctactaagaagaagactcaacgtgtgtgcta 5007
DB 4620 CCTTTTCAGATGGGTAGAGGCTTATCTACTAAGAAAGAGACTTCAACCTGTGTGCTA 4679
QY 5008 agaaatactgaggaanaattttccagaatttgaatgaatacctaaggtiaaagtcagaca 5067
DB 4660 A-AAATATCTGAGGAAATTTTTCGAGATTGGATACCTTAAGGTATTCGGCTCAGACA 4738
QY 5068 atgtcagacttctgtccaggtlaagtlcagagactgtgccaagaatattgggaattgatt 5127
DB 4739 ATGTGCCAGCTTTTGTGCCAGTAATGTAAGGAGCTGGCAAGTATTTGGGATGTATT 4798
QY 5128 ggaacatgcatgtgtacatacagaaccgaagctcagaagcagtgtagagaagtgatagaa 5187
DB 4799 GGAACACTGCATTGTCAATACAGACCCCAAGCTCAGGACAGTGAAGATGAATAAA 4858
QY 5188 ccattaaagaagacccctaccaatgtcacacagagagactgtgattaatgtgtgtgtc 5247
DB 4859 CCATTAAAGAGACCTTACCAATATGACACAGAGACTGCGATTATATATGATAGTTC 4918
QY 5248 tctgtcccttgtgtctttttagagtgtagaacaacccctgagcaagtltggtctgacccct 5307
DB 4919 TCTGTGCCCTTTGTGCTTTTAAAGTTAGAACACCCCTGTGACAGTTTGGCTGACCCCT 4978
QY 5308 atgaattgtcttaagggagaccccccggtgtgacgaagaattgtccttccatagtgtcgt 5367
DB 4979 ATGAATTGCTTACGGGGAGCCCCCGCTGTGTGTAATAATGCTTCTGTACATGTAGTCTG 5038

QY 5368 atgtgtcttcccaagccttcttctctaggtcgaagcgctcaggtgtgtgtgagcagc 5427
DB 5039 ATGTGTGCTTTCCAGCCTCTGTCTGTAGGCTCAAGGCGCTCAGTGGGAGGCAAC 5098
QY 5428 gagcgtggaagcagctccggagagcctactcagg--agggaacttgcaattccaatc 5484
DB 5099 GAGCGTGAAGAGCTCCGGAGGCGCTACTCAGGAGAGGAACTTGGCAAGTTCACATC 5158
QY 5485 gcttccaagtgtgagatcagtcatagttataagcgcacagctgacgaagaacctcgagactc 5544
DB 5159 GCTTCCAAATGGAGATTGATGTATGTATTGAGCGCACCCGCTCAGAGAAACCTCGAGACTC 5218
QY 5545 ggtggaaggagaccttactcgtacttctgacacacccaagcgctgttgaagtgcgaagaa 5604
DB 5219 GGTGGAAGGGCCCTTATCTCGTACTTTGACCACACAAACGCGCTGGAAGTGCAGGAA 5278
QY 5605 tcccacactgtagatcatatcccaagcttaagcgcgcacactcccgacttgggtgtga 5664
DB 5279 TCTCCACTGTGATCCATGATCCACAGTTAAGCTGGCGCCACTCCGACTGGGGTGA 5338
QY 5665 gagcgaagaagacttgagaatcccttgaagcttgcctccatcgctgttcttactcta 5724
DB 5339 GAGCCCAAAAGACTGGAATCCCTTAAGCTTGCCCTCATGCGCTGTCTTACTCTA 5398
QY 5725 acataactcoccagagcagtaglaaacycctatagacagctcgaaaccccatagact 5784
DB 5399 ACAATTAACCTCCAGGCCAGTAGTAACGCTTATAGACAGCTCGAACCCCATAGACT 5458
QY 5785 ttatcccttactcgtctgtattatgacccctgtatacgggtgtcaactgttaattgacctga 5844
DB 5459 TTATCCCTTAACCTGCTGTATTATTAAGCCCTGATACGGGTGTCATGTAATATGACTGCA 5518
QY 5845 ggtgtgtcttccctgaaggaacctgtgtgcctgaaactcatttctgtccatgtattaac 5904
DB 5519 GGTGTGTCTCTAGAGGACACTGTGTGCTGAACTGCATTTCGCTCCGATTGATTAAAC 5578
QY 5905 cccgcgtttaaagaacacactcccaacctagctccgtagtattatggttcttactgtgcga 5964
DB 5579 CCGCGTGTAAAGCAACACTCCCACTAGTCCGATAGTATGAGTTCATGCTGCCCCA 5638
QY 5965 ggcacagaagaagaatactcgtgtgtgtgtcgtgggaatcctctgttagagatgtgagc 6024
DB 5639 GGCACAGAGAAAGAAATTACTGTGGGGTTCGAGGAATCCTTGTGTGAGATGTGAGC 5698
QY 6025 ttgcgtcaactcccaagatgagagactgaaatgtgcgactcctccagaagccggtaaaa 6084
DB 5699 TCGGTCACTCCCAACATGAGAGACTGGAATGGCCGATCTCTCCAGAGCCGGTAAAA 5758
QY 6085 ttctccttltgcaattccggcccggaagtlacaaaatgatlgaactataataagaatag 6144
DB 5759 TTCTCTTGTTCATATCCGGCCCGGGCAAGTACAAAGTGTAACTATATTAAGATTAG 5818
QY 6145 agctgtctcccatcagaacttagattatctaagaatagttactgt--aaaggaacaag 6202
DB 5819 AGCTGTCCCATTCAGACTTATGATTATCTAAAGTTAAAGTTTCACTGAAAAAGAAAAACAG 5878
QY 6203 gaaataatcaaaagtgtataatgtatagctgtggaatagtttttat-tataatggc 6261
DB 5879 GAAATATTTTCAAAAAGTGTGATTAATGTATAGCTGGGGAATAGTTTATTAATATGAGC 5938
QY 6262 gggggagcaggtgtcactttaaccatcgccttagatagtagaagcggggaagaacccctc 6321
DB 5939 GGGGAGACAGGCTTCACTTAACCATTCGCTTGAAGTAAAGACGGGGAGAGAACCCCTC 5998
QY 6322 gtggcaatgggagccgaataagtaactgtgtgaacaagggcccgccctgtgagcagc 6381
DB 5999 GTGGCAATGGGAGCCCGATTAAGTACTGTGTAACAGGGGGCCCGGCGCTGAGAGCCACG 6058
QY 6382 cataactgtccgtgtgcccataaactcgtcgtgagctgtgacataaacaagccgctagc 6441
DB 6059 CATAACTTTCGCGGTGCCCAATTAACCTCGTGGGCTGACATTAACACAGCCGCTAGC 6118
QY 6442 aacagtlacacatgtgatltcctaccacaacagcctagaagaactccccaagtgctcctgtt 6501

Db	1390	ACAACGAGAGAGAGAGAGAAATTTCTTACAGGCTTAAAAAATGTTCTTGGGGCCGAC	1449
QY	1440	ggcgagaccagcgatgcaaaatgagatlgacatggatltcccttgatcgcccggt	1499
Db	1450	GGGCGACCCACCCAGTTCAAAATGATGATGACATGGGATTTCCCTTGACTCCGCCGCT	1509
QY	1500	tgaggctacaacagcgctgagatgagagagcttgaataatctatcgcaagcttgltg	1559
Db	1510	TGGACATCAACACGCGCTGAAAGGTAGGGAGCTTGAATAATCATCCCGAGCTTGCTG	1569
QY	1560	gcgggtcccgggcgccctcaagacggcccaataatttgctgaagtaagaaggtgag	1619
Db	1570	GGGGTCTCCGGGGCGCTCAAGACGGCCCACTATTTGGCTTAAGGTAAAGAGGTGATG	1629
QY	1620	caggagaccgaagcaactccctcgatattcttgagagctcatggaagccttcaagcg	1679
Db	1630	CAGGACCGAAGGAACCTCCCTCGATTTCTTGAGAGCTCATGGAAGCCTTCAGGCGG	1689
QY	1680	ttcaacccttttgatctactacccagagagccagaagaagccttagtgcccttgctcat	1739
Db	1690	TTCAACCCCTTTGATCTTACTCTAGAGAGCCCAAGAAAGCTCACTGGCCCTTGCCCTTCA	1749
QY	1740	ggcgagctcgctctgagatcaggaagaaactcagagactggaaggttacaagagct	1799
Db	1750	GGGCGATCGGCTCTGATATCAGGAAGAACTTCAGAGACTGGAAGGTTTACAGAGCT	1809
QY	1800	gagtiacgtatctagtgagagagagagagaggttatatacagaagggagacagaag	1859
Db	1810	GAGTACGTATCTAGTGAAGAGAGAGAGAGAGGTATTTACAGAGAGAGACAGAAAG	1869
QY	1860	gag	1919
Db	1870	GAG	1929
QY	1920	caag	1979
Db	1930	CAG	1989
QY	1980	ag	2039
Db	1990	AG	2049
QY	2040	ag	2099
Db	2050	AG	2109
QY	2100	aactgcccccaag	2159
Db	2110	AAGTCCCCCAAG	2169
QY	2160	tagggag	2219
Db	2170	TAGGGAG	2229
QY	2220	caacgaagttaggttcttgatgataccgagcggagcattcagtgctgtacaacat	2279
Db	2230	CACACAGTGTGAGTCTTGTTGATACCGGAGCGGACATTCAGTGTCTTACAAACATTA	2289
QY	2280	ggaagaactaaagaagagagagagagagagagagagagagagagagagagag	2339
Db	2290	GGAAGAAGTAAAG	2349
QY	2340	tgagataccgag	2399
Db	2350	TGAGACTACCGAG	2409
QY	2400	atcccttagtgag	2459
Db	2410	ATCCCTTAGTGAG	2469
QY	2460	attctcttgaagaag	2519
Db	2470	ATTCTTTTGAACAG	2529

QY	2520	accctcaattagatgatatcagatatatctcccaagtaaaagcctgataagat	2579
Db	2530	ACCCCTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2589
QY	2580	atacagctccttggttgagagagagagagagagagagagagagagagagagag	2639
Db	2590	ATACAGTCTCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	2649
QY	2640	gcaag	2699
Db	2650	GCAAG	2709
QY	2700	ag	2759
Db	2710	AG	2769
QY	2760	atcccaag	2819
Db	2770	ATCCCAAG	2829
QY	2820	aggaagccttgag	2879
Db	2830	AGGAAGCCTTGAG	2889
QY	2880	gtgag	2939
Db	2890	GTCGAG	2949
QY	2940	gaacgag	2999
Db	2950	GAAAG	3009
QY	3000	cccaag	3059
Db	3010	CCCAAG	3069
QY	3060	caagctaaccttgag	3119
Db	3070	CAGCTACCTTGAG	3129
QY	3120	gccctacacag	3179
Db	3130	GCCCTACACAG	3189
QY	3180	tacgtgagatgag	3239
Db	3190	TACGTGAGATGAG	3249
QY	3240	gcaactactgag	3299
Db	3250	GCACACTAGCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG	3309
QY	3300	tgacgag	3359
Db	3310	TGACGAG	3369
QY	3360	gagcag	3419
Db	3370	GAGCAG	3429
QY	3420	gagtttttgggggag	3479
Db	3430	GAGTTTTTGGGGAG	3489
QY	3480	gccccactaacccgctaaccccaagaagaagaagg	3514
Db	3490	GCCCCACTTAACCCGCTTAACCAAGAAAGAGGGG	3524

RESULT 9
ARI130474
LOCUS ARI130474 7333 bp DNA PAT 16-MAY-2001

[illegible]

Db	2950	GAACGGAACAGGTACACAGTATTGGACTTAAATAAGATGCTTCTTCTGGCTGAGATTACAC	3005
QY	3000	cccaatagccaacacacttttbbcccttgaaatgaggaatccaggtacgggaagaacccgg	3055
Db	3010	CCCACTAGCCAAACACTTTTAACTTTCGAATGGAGAAATCACAAGTCAGGGGAACCCGG	3065
QY	3060	cagctaaccttgaccggaactgcccaaaaggttcaagaatcccgacacacttttaagaa	3115
Db	3070	CACCTCACTGGACCCGAGCTGCCCAAGGGTTCAAGAACTCCCGACACTTTTACGAA	3125
QY	3120	ggccctacaagaggaactcggccaacttaagatccacaacccctcaagttgaacccctccag	3175
Db	3130	GCCTTACACAGGGAGACTTGGCCAACTTAGGATCCAAACACCCCTCAGGTGACCTCTCCAG	3185
QY	3180	tacgttgatagacactctcttcgsgggagaccacaacagagactgcgttgaagaatcagaa	3235
Db	3190	TACGTGATGACCTGCTTCTTGCGCGGAGCCAAACAGAGCTGTTGAAGATCAGAAAG	3245
QY	3240	gcaactactgtcgtgaattgtctgaaactagcctacagaagcctctgctcaagaagccagatt	3295
Db	3250	GCACACTAGCTGGAATTGTCTGACACTAAGGCTAAGAGCCTTGCTTAAGAAAGCCCAAGTT	3305
QY	3300	tcccaagaaagaggttaacatactctcgggtgtacagtttcgsgggcggaagcagatgtcagc	3355
Db	3310	TGCAGAGAGAGGTAAACTCTTGGGGTACAGTTTGGCGGGGACAGCATGGCTGACG	3365
QY	3360	gaggaacggaagaaacttgaftccgaatataccggccccaacaaagccaacaaagtga	3415
Db	3370	GAGGACCGGAAGAAACTGTATGTCAGATACCGGCCCAACACAGCCAAACAAGTGA	3425
QY	3420	gaagttcttgggagacgtcgtatcttgcagactgtgatccggggttgcgaacttga	3475
Db	3430	GAGTTTGGGGAGACGCTGATTTTGCACACGTCGTGATCCCGGGGTTTGCACCTTAGCA	3485
QY	3480	ggcccaactcaaccgcttaacccaagaagaagggg	3514
Db	3490	GCCCACTCTTACCCGCTAACCAAAAGAAAAGGGG	3524

Query	3032	gagagatccaggtacggagaagaacgggcagctcactctgacccgactgccccaaggtt	3091
Db	21	AGATGCTTCTTCTGCTGAGATTCACACCCACATGACCAACACTTTTGCTTCATAG	80
QY	2972	agatgcctctctctgcctgagattacacccactagccaacacttttgcttgcgattg	3031
<p>Query Match</p> <p>Best Local Similarity 100.0%; Pred. NO. 0;</p> <p>Matches 2682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p> <p>32.7%; Score 2682; DB 6; Length 3320;</p>			
<p>BASE COUNT 922 a 844 c 832 g 722 t</p> <p>ORIGIN</p>			
<p>FEATURES</p> <p>SOURCE</p> <p>1. 3320</p> <p>Location/Qualifiers</p> <p>/organism="porcine endogenous retrovirus"</p> <p>/db_xref="taxon:61673"</p>			
<p>REFERENCE</p> <p>AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.</p> <p>TITLE PORCINE RETROVIRUS</p> <p>JOURNAL Patent: WO 9740167-A 1 30-OCT-1997;</p> <p>Q ONE BIOTECH LTD (GB)</p>			
<p>VERSION 1 (bases 1 to 3320)</p> <p>KEYWORDS retroviruses; 1-Mammalian type C virus group.</p> <p>SOURCE porcine endogenous retrovirus.</p> <p>DESCRIPTION porcine endogenous retrovirus.</p> <p>VIRUSES: Retroid viruses; Retroviridae; Mammalian type C</p>			
<p>DEFINITION Sequence 1 from Patent WO9740167.</p> <p>LOCUS A66551</p> <p>ACCESSION A66551</p> <p>VERSION A66551.1</p> <p>KEYWORDS GI:4538104</p>			
<p>RESULT 10</p> <p>A66551 3320 bp DNA PAT 29-MAR-1999</p>			

Db 81 GAGAGATTCAGGATCGGAGAGAACCGGAGACTTACTGAGCCGATGGCCCAAGGTT 140
Qy 3092 caagaactcccgagacactcttgaagaagccctacacaggaacctggccaactcagat 3151
Db 141 CAAAGACTCCCGGACCATCTTTGACGAGGCCCTACACAGGAGACCTGGCCAACTTCAAGAT 200
Qy 3152 ccaacaccttaagtgagccctcccgagtaagtgatgacctgtcttgcggaagcac 3211
Db 201 CCAACACCTCAGGTGAGCCCTCTCCAGTACGTGATGACCTGCTTGTGGGGAGGCAC 260
Qy 3212 caacagactgcttaagaagtaagaagcaactactgctgaattgtcctagacta 3271
Db 261 CAACAGAGACTGCTTAAGAGGTAGAGGACACTACTGCTGGAATTTGCTGACCTAGCCTA 320
Qy 3272 cagagcctctgtaagaagccagatcttcagagagagagtaataactcttggtgacag 3331
Db 321 CAGAGCCTCTGCTAAGAGGCCAGATTTGACAGAGAGAGGTAACTTGGGGTACAG 380
Qy 3332 ttctcgaggcgagagcaactgctgacgagagcaaggaagaactgaagcagatacc 3391
Db 381 TTGCGGGGGCGGACGATGCTGACGAGGACGAGAGAAACTGTAGTCCAGATACC 440
Qy 3392 ggccccaacacagccaacaaagtaagaagtttggggaagctgatttgcagact 3451
Db 441 GGCCCAACACAGGCAACAAAGTAGAGACTTTTGGGACAGCTGATTTGCAGACT 500
Qy 3452 gtgatacccgaggttgcgaccttaagagcccaactctaccgcttaacaaagaag 3511
Db 501 GTGAGTCCGGGGGTTTGGACCTTAGCAGCCCACTACCCGCTAACCAAGAAAGG 560
Qy 3512 gggatctccctgggctcctgagcccaagaagcatctgactcactcaaaaagccctgct 3571
Db 561 GGGATTTCTCTGGCTCCTGAGCACCAAGAGCATTTGATCTTCAAAAAGGCCCTGCT 620
Qy 3572 gagcgacactgctctgacctccctgacgtactaaaccccttaaccttactgtgata 3631
Db 621 GAGCGACCTGCTGCGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGATGA 680
Qy 3632 gctgaagagtagagccgaagagtttaacccaaccccttaagacatgagagaacctgt 3691
Db 681 GCGTAGGGAGTAGGCCAGAGAGTTTAAACCAACCTTAGGACCATGAGAGAGCCGT 740
Qy 3692 tgcctaccctgtcaagaagactgactcctgtagcagtggttggcccgatgtctgaagc 3751
Db 741 TGCCTACTGTCAAGAGCTTGATCTGTAGCCAGTGGTGGCCCGATGTCTGAAGC 800
Qy 3752 tatcgacagctgtgacatactggtcaagaagcgtgaacaattgactttggagcagata 3811
Db 801 TATCGACAGCTGTGCCATACTGTGACAGAGCGCTGACAAATTTGACTTTGGGACAGAAAT 860
Qy 3812 aacttaagccccccctgactgtagaagaactgttggagagcccccaagcagatg 3871
Db 861 AACTGTAAATAGCCCCCATGCACTTGTGAAGACATCGTTCGGCGCCCCAGACCGATGGAT 920
Qy 3872 gaccaagcccccgatgacccaactcaagaacctgtcttcaagaagaggtcaacttgc 3931
Db 921 GACCAAGCGCCGATGACCCACTATCAAAAGCCTGCTTCTCACAGAGAGGTACATTTGGC 980
Qy 3932 tccacacagcgtctcaacactgcaactcttctgcctgaagaagactgtaacagtgac 3991
Db 981 TCCACAGCGCGCTTCACACCTGCGCACTTCTGCTGAGAGAGATGTAACAGTGAAC 1040
Qy 3992 taatattgccaactaacttgaattgagagagctgggtcccaagagcccttaagat 4051
Db 1041 TATGTGTTGCCATCAACTATTGATTGAGAGAGACTGGGGTCCCAAGAGACTTTACAGAT 1100
Qy 4052 accgctgactggaagagtgtaactggttcaactgacggaagcagctatgtgtggaag 4111
Db 1101 ACCGCTGACTGGAGAGTGTCTAACTGCTTACTGACGGAACAGCTATGTGTGAAGG 1160
Qy 4112 taagagatgctggggcgagtggtgtgagcgagccgcaagatctggtgcagagact 4171
Db 1161 TAAGAGATGCTGTGGCGGCAGTGTGAGCGGAGCCCGACAGATCTGGGCGCAGACCT 1220
Qy 4172 gccggaaggaacttaagcgaagaagctgagctcaatgagccctcaagcagcttgcgct 4231
Db 1221 GCCGGAAGGAACCTTCAAGCCCAAAAGGCTGAGCTCATGGCCCTCACCAAGCTTTGGGCT 1280
Qy 4232 gccgaaggaatccataaacttatacggacagcagtgatgacctgtgacatgcaca 4291
Db 1281 GGCCGAAGGAATTCATTAACATTTATACGACACAGCTATGAGCTTTGCACGTGCACA 1340
Qy 4292 cgtacaagggccaactataaanaaggggtgtcttacctgaagcgggaaggaataa 4351
Db 1341 CGTACACGGGGCCATCTATTAACAAAGGGGGTGTCTTACCTCAAGCGGAGGAATTA 1400
Qy 4352 gaacaaaggaatcttaagcctatagaagccttactatcttgcocaaaaagctagctat 4411
Db 1401 GAACAAAGAGAAATTTCTAAGCTATTAGAAGCTTACATTTGCCAAAAAGCTACTAT 1460
Qy 4412 tatacactgtctgacatcaagaagccaagaatctcatactataagagggaaccatgagc 4471
Db 1461 TATACACTGCTGAGACATCAGAAACCAAAAGATCTCATATCTAGAGGAGACCAATGCG 1520
Qy 4472 tgaacgggttgccaagcagcccaagcgtgttaaccttgcctataatagaacgac 4531
Db 1521 TGACCGGGTTCAGAGCGAGCGAGCCAGGCTGTTAACTTTGCTTATTAACAAACGCC 1580
Qy 4532 caaagccccagaacccaagacagtaacacctagaagactggcaagagataaaaaagat 4591
Db 1581 CAAAGCCCCAGAACCCAGACAGACAGTACACCTTAGAAGACTGCGAAGAGATTAAGAAAT 1640
Qy 4592 agaacgtctctgagaactccggaaggggacctctactactataggggaagaactct 4651
Db 1641 AGACCACTTCTGAGACTGCGAGGGGAGGACTCTTATACCTCATATGTGGAAGAAATCT 1700
Qy 4652 gccccaagaagaaggttagaataatgtccaagaataactcgtctaaaccaactagaac 4711
Db 1701 GCCCCAGAAAGAGGGTTAATAATATGTCCAAAGATATCATGCTTAACCCACTAGAGAC 1760
Qy 4712 taaacactgtcagcagtggttcagaacatcccttactcaatgtctgaggtaccagagt 4771
Db 1761 TAAACACTCTGACAGTGTGTCAGAAACATCCCTTATCATATGTTGAGGCTACAGAGAT 1820
Qy 4772 ggtgactcgtgtgtcaaacatgtgtgtccctgcagcgtgtgtaagtaagtaactccag 4831
Db 1821 GGCTGACTGCTGTGTAACATTTGTGCTCCCAAGCTGTTAATGCTTAATCTTCAG 1880
Qy 4832 aatacctcagaagaagacttaaggggaagcccaagcgctcaactggaagtgagct 4891
Db 1881 AATACCTCCAGGAAGAGACTTAAGGGAAGCCACCGGCGTCACTGGGAGATGGAGCT 1940
Qy 4892 cactgaagtaagccgctaaatacgaagaacataactatgtgttctgtaaacctt 4951
Db 1941 CACTGAGGTAAAGCGGCTTAATAGGGAACAAATATCTATTTGTTTGTAGACACTT 2000
Qy 4952 ttcaagatggtagagcttactactaagaagaagacttcaacgltgtgactaagaa 5011
Db 2001 TTCAAGATGGGTAGAGCTTATCTACTTAAGAAAGAACTTCAACCTGTGTGGCTAAGAA 2060
Qy 5012 aatactgagaagaatttccaagatttgtaatacctcaagtaagtaagtgcaagaatg 5071
Db 2061 AATACCTGAGAGAAATTTTCCAAATTTGGAATATCTAAGGTATAGGATCAGACATAGG 2120
Qy 5072 tccagactgtgtgccagtgtaagtcaggaactggtcccaagatattgggattgtgaa 5131
Db 2121 TCCAGCTTGTGCTCCAGGTAATTCAGGAGATGCGCAAAATATTTGGGGATTTAGTGGAA 2180
Qy 5132 actgcatgtgtcataagagcccaagcctgaagcaggtagagaagtaagaacat 5191
Db 2181 ACTCATTTGTGCATACAGACCCCAAGCTCAAGCAAGTGAAGGATGATGAACAT 2240
Qy 5192 taaagagaccttaacaaatgacacagagactggcatlaatgatagtgactcct 5251
Db 2241 TAAAGAGACCTTACCAAAATTTGACACAGAGACTGGCATTAATGATTTGATGGCTCTCT 2300

Db 1165 CATTTCGATTTGGATGGCCATCAGAGGGAGCCCTTATTCCTGAATATTCCTGGCTGTA 1224
 Qy 781 aggaatcatcttctagactggagccggctcatctcatctagagccctatctcta 840
 Db 1225 AGGCAATCATTTTTCAGACTGGACCGGCTCATCTCATGAGAGCCCTATATCTTA 1284
 Qy 841 cgtggcaagaatttggcagaagaatctccgcacatgggttaaacaatgctaaataaaca 900
 Db 1285 CGTGGCAAGATTGGCAGAGATCTCCGCCATGGGTAAACCATGGCTAAATAAACCAA 1344
 Qy 901 gaaagccaggttcccggaatctctgctcttggagaagaataaacaactcgccggaagaag 960
 Db 1345 GAAAGCCAGGTCCCGGATCTGCTCTTGGAGAAGAAACAACACTCGGCCGCAAAAG 1404
 Qy 961 tcgagccctctctctgctatctaccccgagatcgaggagccgcgacttggccggaacc 1020
 Db 1405 TCGAGCCCTC-TCTCTGATCTACCCGAGATCGAGAGCCGCGCACTTGGCGGAACCC 1463
 Qy 1021 caactgttcccccacccctatccagcaagggtgctgtgaggga-ctctgcacct 1079
 Db 1464 CAACCTTTCCCCACCCCTTATCCAGACAGGTCCTGTGAGGGGACCTCTGCCCT 1523
 Qy 1080 cctggaactccggtgtgtggaggagccgtctgcccggagactggagccggaaggcgaccc 1139
 Db 1524 CTTGAGCTCCGGTGTGGAGGACTCTGCTCCGGGACTGGAGCCGGAAGAGCGCCACC 1583
 Qy 1140 ccgagagcggaacagagatcgcatatlaacgctgcacactatggccctcccatgca 1199
 Db 1584 CCGAGGGGAGACAGAGATGCGATATTACCGCTGGCGACCTATGCCCTCCCATCCG 1643
 Qy 1200 gggggccaattgcaagccctccaglatgtgagcccttctctggaagctctaatatgg 1259
 Db 1644 GGGGGCAATTTGACGCCCTCCAGATTTGGCCCTTTCTCTGAGATCTATTAATTGG 1703
 Qy 1260 aaactaaccatcccccctctctcgagagatccccaacgctcgaagggttggtagtcc 1319
 Db 1704 AAACTTAACTATCCCTCTTCTCGAGAGATCCCAAGCCCTCAGGGGTTGTGAGATCC 1763
 Qy 1320 ctatgtctctcaccagactcttggagatgtcacaagctgtctgcagacactctc 1379
 Db 1764 CTTATGTTCTCTCACACACTACTTGGGATGATGTCTGACAGCTGTGCACACTCTTC 1823
 Qy 1380 acaacccggagagcgagagaatctcttgaagctgaagaaatgttccctggaggccagc 1439
 Db 1824 ACAACCGAGAGCGAGAGATTTCTGTAAAGCTTAAGAAATGTTCTCTGGGGCCGAC 1883
 Qy 1440 ggggacccacgcaatgtcaaatgagatgacatggagatttcccttgaactgcgcccgt 1499
 Db 1884 GGGCCACCCACGCACTTGTCAAAATGAGATGACATGGGATTTCCCTTACCTGCCCGGT 1943
 Qy 1500 tgggaactaacaacagcgctgaaggttaggagagcttgaatctatcgccagctctgtg 1559
 Db 1944 TGGGACTTACAAACACGCTGAAGGTAGGAGAGCTTGAATAATCTACGCCAGGCTCTGTG 2003
 Qy 1560 gggggtctccggggcgccctcaagacgagcccaactaatgttgctaaagtgaagagtga 1619
 Db 2004 GGGGGTCTCCGGGGCGCTCAAGAGAGCCCACTTAATTGGCTAAGAGTAGAGGTATG 2063
 Qy 1620 caggagacgaacgaactccctcgttattcttggagagctcaatgaaagccttcaagcgg 1679
 Db 2064 CAGGGACCGAAGAACCTCCCTCGGTATTTCTTGAAGAGCTCATGGAAGCCCTTCAGGGCG 2123
 Qy 1680 ttaaccccttctgactctcctcaagagcgccgaagagctgagtgcccttgccctcat 1739
 Db 2124 TTTCACCCCTTTTGTATCTTAACCTCGAGGGCCGAGAAAGCTTCAGTGGCCCTTGCTTCATT 2183
 Qy 1740 gggcaactcgactctgatactaggaagaactcgaagacttgaagggttlaaaggagct 1799
 Db 2184 GGGCACTCGCTCTGATATCAAGAAAGAACTTCAGAGACTGGAAGGCTTACAGAGAGCT 2243
 Qy 1800 gaagtaactgactagtagagagcgagagaggtgtatlaaagaaggagacagaagag 1859

Db 2244 GAGTTACGTGATCTAGTAGAGAGGACAGAGAAGTGTATTACAGAAAGGACAGAAAG 2303
 Qy 1860 gagaaggaacgagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1919
 Db 2304 GAGAAGGAACAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2363
 Qy 1920 caagaagaatcttactaagaatcttggccgagtggttgaagggaagaagaagaaga 1979
 Db 2364 CAAGAGAAAGATTGATTAAGATCTTGGCCGAGTGTGAAGGAGAGAGAGAGAGAGAG 2423
 Qy 1980 agagagagagatcttaagaagaatgtgtcagagcccttaacagctgaaggagagag 2039
 Db 2424 AGAGAGAGAGATTATAGAAATTAAGTGTAGGCTCTAGACACTGAGAGAGAGAGAG 2483
 Qy 2040 agagacccacactgcaagaagcagtggtgtatcttgaagaagaagaagaagaaga 2099
 Db 2484 AGAGCCCACTGTGACAG 2543
 Qy 2100 aactgcccagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2159
 Db 2544 AACTGCCCAAG 2603
 Qy 2160 taaggagagcgggtgttcgagcccccctcccgagccagagtgtaacttgaagtgagg 2219
 Db 2604 TAAGGAGAGAGGGGTTGAGACCCCTCCCGAGCCAGGCTAACCTTTGAAGAGAGAGAG 2663
 Qy 2220 caaccagtgagctctgtgtatgataccggagcggagacttcaagctgtctacacat 2279
 Db 2664 CAACCGATTGAGTCTCTGTTGATACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2723
 Qy 2280 ggaanaactaaagaanaaataatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2339
 Db 2724 GGAATACTAAAGAAAAAATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2783
 Qy 2340 tggactaacccgaagaacggttgaacttggaggtggaacgggtgaaccactgttctgtc 2399
 Db 2784 TGGACTAACCCGAGAGAACCTGTGACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2843
 Qy 2400 atccctgagtgcccaatccctctctaggttagagacttcaagcaagaatggagactca 2459
 Db 2844 ATCCCTGAGTCCCGAGTACCCCTTCTAGTAGAGACTTACTGACCAAGATGGAGCTCAA 2903
 Qy 2460 atttcttgaacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2519
 Db 2904 ATTCTTTTGAACAG 2963
 Qy 2520 accctcaatagatgatactgatactatcttccccaagtaagcctgtatcaagat 2579
 Db 2964 ACCCTCAATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3023
 Qy 2580 at 2581
 Db 3024 AT 3025

RESULT 12
 AF014162
 LOCUS 1971 bp mRNA VRL 01-JUN-1998
 DEFINITION Porcine endogenous retrovirus envelope protein (env) mRNA, complete cds.
 ACCESSION AF014162
 VERSION AF014162.1 GI:3169735
 KEYWORDS
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 porcine endogenous retrovirus
 Viruses; Retroid viruses; Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 REFERENCE
 AUTHORS Hamworth, C., Galbraith, D.N., Lees, G.M. and Smith, K.T.
 TITLE Porcine endogenous retrovirus (PoEV) Env sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1971)
 AUTHORS Hamworth, C., Galbraith, D.N., Lees, G.M. and Smith, K.T.

|||||
Db 1621 GAAGGCGGCTTATGTCAGCCTTAAAGAAATGTTGCTTCTATCTAGATCACTACAGCA 1680
Qy 7300 gccataagagcctccatgaaacagcttagaaaaaagtagagagcgctgaaggaaga 7359
Db 1681 GGCATGAGACATCCATGAAACAGCTTAGAAAAAGTTAGAGGCGGTGAAGGGAAGA 1740
Qy 7360 gaagctgacacaggggtgtttgaagatgttcaacagctctcctctgtagaacaccg 7419
Db 1741 GAGGCTGACCAAGGGGTGTTGAAGATGTTCAACAGCTCTCTTGATGACCACTCG 1800
Qy 7420 cttctgctcgaacggggccctagtagctcgtctcttacttaacgttggcctgc 7479
Db 1801 CTTTCTGCTGACGGGGCCCTAGTAGTCTCTCTCTGTTACTTACAGTTGGCCTTGC 1860
Qy 7480 ttatttaagatgttctgctcttctgtagaagacagtagagtagtccagatcagta 7539
Db 1861 TTAAATTAATGTTGTTGCTTGTGTTAGGAACGAGTAGTCACTCAAGATCATGTGA 1920
Qy 7540 cttaagcaacagtagcaagcctctcgaagcaagagaagtagtactctag 7590
Db 1921 CTTAGGCAACAGTACCAAGGCTTGTGAGCAAGAGAAACTGACTCTAG 1971

RESULT 13
LOCUS A66559 1974 bp DNA PAT 29-MAR-1999
A66559
DEFINITION Sequence 9 from Patent WO9740167.
ACCESSION A66559
VERSION A66559.1 GI:4538112
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 1974)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 9 30-OCT-1997;
Q ONE BIOTECH LTD (GB)

FEATURES
Source Location/Qualifiers
1..1974

BASE COUNT 534 a 489 c 483 g 468 t
ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 18.3%; Score 1499; DB 6; Length 1974;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

Qy 5639 ggagcaaccccgactcgggtgtgagagcgaagaagactgaatcccttaagttc 5698
Db 20 GGGCCACCTCCGACCTGGGGTGAGAGCCGAAAGACTGGAATCCCTTAAAGCTTCG 79
Qy 5699 cctccacgcctgttccctactctaaacaaactccccaagcagtagtaaacgcctta 5758
Db 80 CCTCCATCGGCTGGTTCCTTACTCTAACAATACTCCCGCAGGCAAGTAAGAAACGCTTA 139
Qy 5759 tagacagctgaacccccaatagaccttacccttaccctgctgattatgacctgata 5818
Db 140 TAGACACTCGAACCCCAATAGACCTTATCCCTTACCTGCGATATTATGACCTGTATA 199
Qy 5819 cgggtgtcactgttaaatagacacagagttgtctccttagagcactgtgtgcgaac 5878
Db 200 CGGGTGTCACTGTAATAAGACCTGAGGTGTGCTCTAGAGGACCTGGTGGCCTGAAC 259
Qy 5879 tgcattctgctcctcgaattgattaaacccgcgtttaaagaacaacccccaactagtc 5938
Db 260 TGCATTTCTGCGCTCGATTGATTAAACCCGCTGTAAAGACACACCTCCCAACCTAGTC 319
Qy 5939 gtagttatgggtcttatctgtcgcacagacagagaagaataactgtgggggtctc 5998
|||||

Db 320 GTAGTTATGGGTTCTATTGCTGCCCCAGCACAGAGAAAGAAATACTGTGGGGTCTG 379
Qy 5999 ggaatccctctgtagagatgagctgtgctacccccaagatgagatgagatgagc 6058
Db 380 GGGATTCCTTCTGTAGAGATGAGACTCGTCCACTCCAAAGATGAGACTGGAATGTC 439
Qy 6059 cgaatcctcccaagaccgggtlaaaattcccttctgctcaattccggccgggcaagta 6118
Db 440 CGATCTCTCTCCAGAGCCGGGTAATAATCTCTCTTGTCAATTCGCGCCGGGCAAGTACA 499
Qy 6119 aatgataaactatataaagaatgactctccccaatagacttagatattataaaga 6178
Db 500 AATATGATAAATATATTAAGATTAAGACTGCTCCCACTCAAGACTTAATTAATTAAGA 559
Qy 6179 taagttcaactg-aaagaaacaggaataattcaaaagtgtgataaagtgtagact 6236
Db 560 TAAAGTTCACTGTAAGAAAGAAACAGAAATATTCAAAAGTGAATATGATGACT 619
Qy 6237 ggggaatagttttat-tatatggcgggggaagglccaacttaacatctgcctta 6295
Db 620 GGGGAATAGTTTATTAATATGCGGGGGAGCAGGGTCCACTTAAACATTCGCCCTTA 679
Qy 6296 gtagtagagacgggagacgaacccctgtgtgaaatgtgagaccgataaagtagctgagc 6355
Db 680 GGATGAGACGGGGACAGAACCCCTGTGCAATGGACCGATTAAGTACTGGCTGAC 739
Qy 6356 aagggcccccggcctctgagaccacgcatatctgcgtgtccccaattaaactcgtgc 6415
Db 740 AAGGGCCGCCCGCCTGAGAGCCAGCATTAATTCGCGGTGCCCAATTAACCTGCTGC 799
Qy 6416 ggccctgacataacacagccgcttagaacaagtagtactgtgatttcttcccaacagc 6475
Db 800 GGCCTGACATAACACAGCCGCTTACCAACAGTACGATGATGATTCCTTCAACACAGC 859
Qy 6476 ctagaacatccccagtgcttctgttgaagacagacagactctcagttccatcagg 6535
Db 860 CTAGAAATCCCCAGAGTGTCTGTTAAGACGGGACAGAGACTCTTCAAGTCAATCCAGG 919
Qy 6536 ggccttcccaagcatalcaactccacagccctgtagacactctctctgtgtgtctgc 6595
Db 920 GAGCTTTCACACCATCACTCACTCCAGACCGCATGATGCCACTTCTCTGTGGGCTTGTGC 979
Qy 6596 tatccatagggcctccctattatgagggatgtgtctaaagaaatcaatcagttagcca 6655
Db 980 TATCCCAAGGCGCTCTTATATGAGGGGATGCTTAAAGAAATTCATATGATGACCA 1039
Qy 6656 aagagcatagaataaagttagatagtggtgtccgaataagttacccctcactgaagtt 6715
Db 1040 AAGAGCATAGAAATCAATGTACATGGGGGTCCGAAATTAAGCTTAACCTCACTGAAGTT 1099
Qy 6716 ccgggaaggagacatgacataggaagaagctccccaatcccaacaacacttgcatagla 6775
Db 1100 CCGGGAAGGGGACATGACATAGGAAGAAAGCTCCCATCCACCAACACCTTGTCTATAGA 1159
Qy 6776 ctgtgttataggaagccctcagaanaatcaglattagtagtctgtgttataaagtgct 6835
Db 1160 CTGTGTTTATGAGAGGCTCAGAAATCAATATTATGACTGTTATTAACAGTGTGT 1219
Qy 6836 gggcatgcaatctgtttaaaccctgtgttccacactcagcttcaaccaatcaag 6895
Db 1220 GGGCATGCAATCTGTGGTTAACTCCCTGTGTTTCACTCACTTTCATCAACATCAAG 1279
Qy 6896 attctgtctcctgctcccaatcgtcccgagtgtagtaccatccctgaagaagtgtcc 6955
Db 1280 ATTGTGTGTCAATGTCGTCGCAATCGTCCCGAGTGTACTACCATCTCGGAGTGTCC 1339
Qy 6956 ttgataatatagactatcgtgtataacagcaacaaagagaacgtatcccttaccctag 7015
Db 1340 TTGATGAATATGACTATCGGTATTAACCGACCAAAAGAGAAACCCGTATCCCTACCTAG 1399
Qy 7016 ctgtatgtcctgattagagagccgttgtgtgtatgagaaacagagacagctgcctgata 7075
Db 1400 CTGTATGTCTCGAATTAGGAGCGCGCTTGGCTAGGAACAGGACAGCTGCCCTGATCA 1459
|||||

QY 7076 cagagacacagagctagagaaagacttgtagctacatgcygcacatgacagaatc 7135
|||||
Db 1460 CAGGACACAGCAGCTAGAGAAAGACTTGAGCTACATGCGGCATGACAGAAATC 1519
QY 7136 tccgagcctlaaaggagctgttagcaacctagaagagctccctgctcttgccttag 7195
|||||
Db 1520 TCCGAGCCTTAAGAGAGCTGTTTACCAACCTAGAAAGACTCCCTGACTCTTTGCTGAG 1579
QY 7196 tggctctacagaacccggagagatagatctgctgttcttaagaagaagtgatg 7255
|||||
Db 1580 TGGTCTACAGAACCGGAGGAGATAGATCTGCTTCTTAAGAAAGTGGGTATG 1639
QY 7256 cagccttaaaagaagaagtgctctatgtagatcactcagagccatcagaatcca 7315
|||||
Db 1640 CAGCCTTAAGAAAGAAAGTGTCTTCTATGTAGATCACTCAGAGCCATCAGAACTCA 1699
QY 7316 tgaacaagcctlaaagaagaagtgagagcgctcgaagaagaagaagcctaccagag 7375
|||||
Db 1700 TGAACAAGCTTAGAAGAAAGTTAGAGAGCGCTCGAAGGAAAGAGAGCTACCAAGGCT 1759
QY 7376 ggttgaagagatgltcaacaggtctccttgatgacacccctgcttctgctgagc 7435
|||||
Db 1760 GGTGAAGAGATGTTCAACAGGCTCTCTGATGACCAACCTGCTTCTGCTGAGCG 1819
QY 7436 ggccttagtgctcctgctctgcttactactacagttggccttgccttaatagt 7495
|||||
Db 1820 GGCCTTAGTGTCT 1879
QY 7496 ttgcttcttagagaagaagatgagtgagtcacagatagtgacttagagaagaag 7555
|||||
Db 1880 TTGCTTTGTTAGAGAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1939
QY 7556 aagcctctgagcgaagaagaacatgacctctag 7590
|||||
Db 1940 AAGGCTTCTGAGCAAGAGAACTGACTCTAG 1974
RESULT 14
LOCUS AX002804 3482 bp DNA PAT 21-AUG-2000
DEFINITION Sequence 3 from Patent WO9853104.
ACCESSION AX002804
VERSION AX002804.1 GI:9885132
KEYWORDS
SOURCE porcine endogenous retrovirus.
ORGANISM porcine endogenous retrovirus
REFERENCE 1 (bases 1 to 3482)
AUTHORS Stoye,J.P. and Weiss,R.A.
TITLE Detection of retroviral subtypes based upon envelope specific
sequences
JOURNAL Patent: WO 9853104-A 3 26-NOV-1998;
MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB)
FEATURES
source
1.3482
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"
BASE COUNT 927 a 854 c 867 g 834 t
ORIGIN

Query Match 17.9%; Score 1466; DB 6; Length 3482;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 5639 gggcgacccctccgagctcgggtgagagcgaagaagactgagaatcccttaagctcg 5698
|||||
Db 930 GGGCGACCTCCGACCTGGGGGTGGAGAGCCGAAAGACTGAGAAATCCCTTAAGCTTCG 989
QY 5699 cctccacgcctgttctactactaacaataactcccgagcgaagtagaagaacgctta 5758
|||||

Db 990 CTTCAATCGCTGCTTCTTACTTAACAAATTAATCCCAAGCCAGTAGTAACGCTTA 1049
QY 5759 taagacgctcgaaccccccatagacaccttacccttaccctgctgattatgacccgata 5818
|||||
Db 1050 TAGACAGCTGAAACCCCATAGACCTTATTCCTTACCTGCGATATTATGACCTGATA 1109
QY 5819 cgggtgtcactgttaataagcactcagagtgctgctccttagaaggacactggtgcctgaac 5878
|||||
Db 1110 CGGGTCTCACTGTAAATAGACACTGAGGTGTTCTCTAGAGCACCCTGGTGGCTGAAC 1169
QY 5879 tgcatcttcctcccgatattaaacccgctttaaagaacacccctccaaactgcttc 5938
|||||
Db 1170 TGCATTTCTCTCCCTCCGATTATTAACCCGCTGTAAAGCACACCTCCCAACTAGTGC 1229
QY 5939 gtagtlatgggttctatcttgctccagcgacagaagaagaataactgtgaggttcg 5998
|||||
Db 1230 GTAGTTATGGGTTCTATTGCTGCGCAGGCAACAGAAAGAAATACTGTGGGGTTCTG 1289
QY 5999 ggggaatcctctgttagagatgagagctcgtaacccctcaacagatgagactggaatggc 6058
|||||
Db 1290 GGGAAATCCTTCTGTAGAGATGAGAGCTGCTCACTCACTCAACATGAGAGACTGAAATGCG 1349
QY 6059 cgatctctccagagccgggttaaattctccttctgaattccggccggcgagataa 6118
|||||
Db 1350 GATCTCTCTCAGAGACGGGTAAATTTCTCTTTCTCAATTCGCGCGCAACTTACA 1409
QY 6119 aaatgataaactalaaagataagagctgctcccaacgaacttagatctaaaga 6178
|||||
Db 1410 AAGTGAATGAACATATATTAAGATTAAGAGCTGCTCCCATCAGACTTAGATTATTAAGA 1469
QY 6179 taagtttactg--aaagaaacaggaataatccaagaatgataatgataatgagct 6236
|||||
Db 1470 TAACTTCACTGAAAAAGGAAAAACAGAAATATTCAAAAAGTGATGATGAGCT 1529
QY 6237 ggggaatagtttttat-tatatgggggggagcgaggtccacttaacactgacctta 6295
|||||
Db 1530 GGGGAATAGTTTTTTTATTAATATGCGGGGAGACAGTTCACCTTAACCAATTCGCTTA 1589
QY 6296 ggaatagaacggagacagacacccctgtgcaatgagacccgaataagtaagctgagac 6355
|||||
Db 1590 GGAATAGAGAGGGGAGACAGAAACCCCTGTGGAGTGGAGACCCGATAAGTACTGGCTGAAC 1649
QY 6356 aagggcccccggcctgtgagcgaacgataactgtgcggtgcccgaataacactgcgtgc 6415
|||||
Db 1650 AGGGGCCCGCGGCTGTGAGACACCGCATTAACCTGGCGGCCCAATTAACCTGCTGAC 1709
QY 6416 ggcctacataacacggcgcccttagaagaagtaacactgagatgatttcttaccacaagc 6475
|||||
Db 1710 GGCCTACATTAACACAGCGCGCTAGCACAGTACCACTGATTTGATTCTTACCAACAGCG 1769
QY 6476 ctgaaacctcccaagtgcttctgtttaagaacagagaagactcttcaagctcaccag 6535
|||||
Db 1770 CTAGAAGCTCCCGAGGTGTTCTGTTAGACAGAGACAGACTCTTCAGTCTCANTCCAGG 1829
QY 6536 gagcttccaagccatcaactccacccgacccctgagcactctctctgttgccttgc 6595
|||||
Db 1830 GAGCTTCCAAAGCATCACTCCACGACCCGTGATGCCACTTCTTGTGGCTTGTGC 1889
QY 6596 tatccctcagggcctccttattatgagggagatgctcctaaagaagaataatcaatgtacca 6655
|||||
Db 1890 TATCTCAAGGGCTCTTATATGAGGGGATGGCTAAAGAAAGAAATTCATGATGACCA 1949
QY 6656 aagacataagaataatgatacaggggtcccggaataaagcttaacccctcagaagtt 6715
|||||
Db 1950 AAGAGCATTAAGAAATATATGTACATGAGGGTCCCGAAATAGCTTACCTCACTGAGATT 2009
QY 6716 ccgggaagggagacatgataaggaagctccccaatcccaacacacttgcctatagta 6775
|||||
Db 2010 CCGGGAAGGGAGCATATGATAGAAAGCTCCCGCATCCCAACACACTTTGCTATATA 2069
QY 6776 ctgtggttatgagcagggcctcagaataatcaglatattgactcgtgttatacaggtgtc 6835
|||||
Db 2070 CTGTGTTTATGAGAGGCGCTCAGAAATCAAGTATTTAGTACCTGTTTAAACAGTGTGT 2129

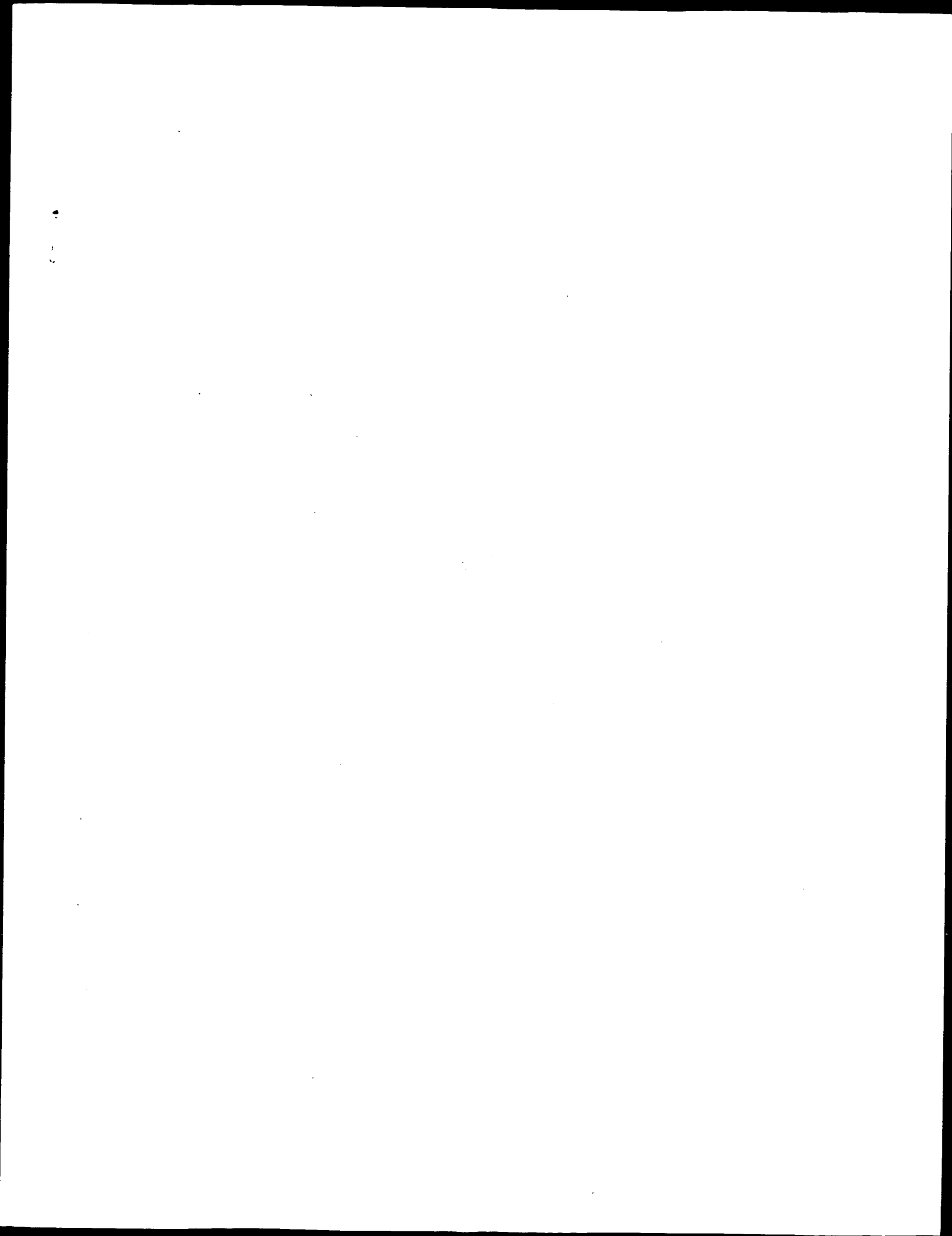
Query Match 17.9% Score 1466: DB 14; Length 3482;
Best Local Similarity 99.3%: Pred. No. 0;
Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 5639 ggcgcacactccgcagctcgaggctgagagccgaagaaagctgagatcccttaagctcg 5638
|||||
Db 930 GCGGCGACCTCCGACCTGGGGTGGAGACCGGAAAGCTGAGAAATCCCTTAAGCTTCG 989
5699 cctccatcgctgltccttactcttaacaataactcccaagccagtaglaaagcctta 5758
|||||
Db 990 CCTCATGCGCTGGTTCCTTACTTAACATAACTCCAGGACAGTAGTAGTAAGCGCTTA 1049
QY 5759 tagaagctgaaccccatagacacttatccctactgctgctatattgaacctata 5818
|||||
Db 1030 TAGACAGCTCGAACCCCATGACCTTATCCCTTACCTGAGCTATTAATTAACCTGATA 1109
QY 5819 cgggtgtactgtataatagcaactcgaggctgtgctccctagaagagacccgtgtgac 5878
|||||
Db 1110 GGGGTGTCACTGTAATATGACACTGAGGTGTGCTCCCTAGAGGCACTGGTGGCTGAC 1169
QY 5879 tgcattctgcctccgattgataaccgcgtgttaaaagacacactcccaactagtc 5938
|||||
Db 1170 TGCATTCTGCTCCGATTTGATTAACCCGCTGTTAAAGCACACCTCCCACTAGTTC 1229
QY 5939 gtagttatgggtctatattgctgcccagagcaagagaaagaaatctgtggggtctg 5998
|||||
Db 1230 GTAGTTATGGGTCTATTGCTGCCCCAGGACAGAAAGAAATATCTGGGGGTTCTG 1289
QY 5999 ggaatccctctgtagagatgagctgagctcaactccaaagatgagactggaatgac 6058
|||||
Db 1290 GGAATCCCTTGTAGAGAGATGAGCTGCGTCACTCCAAAGATGAGAGACTGGAAATGCG 1349
QY 6059 cgatctctctccagagacgggtaaatctctctgtcaatctccgcccgggcaagtaca 6118
|||||
Db 1350 CGATCTCTCTCCAGACCGGGTAAATTTCTTTGCAATTCGGGCGGCAAGTACA 1409
QY 6119 aaatgaagaaactataaagataaagctgctcccatagactatgactataagaa 6178
|||||
Db 1410 AAGGATGAAACTATATTAAGATATAGAGCTGCTCCCATAGACTTAAGATTATCTAAGA 1469
QY 6179 taagtctactg--aaagaaacaggaataatcaaaagtgtatagatgact 6236
|||||
Db 1470 TAACTTTCACCTGAAAAAGAAAAACAGSAAATNTCAAAAGTGTATATGATATGACT 1529
QY 6237 ggggaataagttttat-tatataggcggggagcaaggtccactttaaccatcgctta 6295
|||||
Db 1530 GGGGAATAGTTTTTATTAATATGCGGGGAGAGAGGCTCACTTAACCATTCGCTTA 1589
QY 6296 ggaatagagacgggagacaaacccctgtggcaatggaacccgaataagtaactgtgac 6355
|||||
Db 1590 GGATAGAGAGCGGGAGACGAAACCCCTGTGGAGTGGAGCCGATTAAGATAGCTGAGAC 1649
QY 6356 agggggcccgccgagcagcagacataactgtgcgtgtgcccataaactcgctgc 6415
|||||
Db 1650 AGGGGCCCCGGCCCTGAGACCCGATTAACCTTGGCGGTGCCCAATTAACCTCGCTGC 1709
QY 6416 ggcctgaacataaagcgccctagcaacagtaactgtgattgtattcctacaacagc 6475
|||||
Db 1710 GGGCTGACATTAACACACCCGCTAGAACGCTACACTGATGATTCTTACCAACACCGC 1769
QY 6476 ctagaacactcccaagtggtctgtttaaagcagagagagactcttcagttcataag 6535
|||||
Db 1770 CTAGAAACTCCCAAGGTTCTCTGTTAAGACAGAGACAGAACTCTTACAGCTCATCCAG 1829
QY 6536 ggaacttcaagcatcaactccacagacactgtacactctctctctgtgtgctgtgc 6595
|||||
Db 1830 GAGCTTTCACAGCATCAACTCACCGACCGCTGATGCCACTTCTTGTGGCTTTGTC 1889
QY 6596 taccctcagggctccttattatgagggatggtctaaagaaagaaatcaatgtgacca 6655
|||||
Db 1890 TATCTCAAGGCTCTTATTATGAGGGGATGCTAAAGAGAAATTAATGTGACCA 1949
QY 6656 aagagcatagaatcaatgtatcattgggtlcccgaaataagcttaaccctcaactgaatt 6715

|||||
Db 1950 AAGGACATAGAAATCAATGATACATGGGGTCCCGAAATAAGCTTACCTCAGTAGATT 2009
QY 6716 ccgggaagggagacatgtatggaagaaatctcccaatcccaacaaactttgctaagta 6775
|||||
Db 2010 CCGGGAAGGGAGACTGATAGGAAAGCTCCCATCCACCAACCTTGTGTTAATAGTA 2069
QY 6776 ctgtgttctagaagagcctcagaataatcaatgattagttactgtgttaacagtggt 6835
|||||
Db 2070 CTGTGTTTATGAGCAGAGGCTCAGAAATCAGATTTATGACTGGTTTAAACAGTGCT 2129
QY 6836 gggcatgcaactactgggttaacccctgtgttccactagtlcttcaacaaatccaaag 6895
|||||
Db 2130 GGGCATGCAATTAATCTGGTTTAAACCCCTGTGTTCCACTCAGCTTCAACCAATCCAAAG 2189
QY 6896 attctgtgtacatggtcccaatgtcccccagagtgtaacacatccttgaggaagtgtcc 6955
|||||
Db 2190 ATTCTGTGTCAATGTGCTCAATGCTCCCGAGTGTACATCCATCTGAGAAAGTGTCC 2249
QY 6956 ttgatgaatatagactatcgatataccgaccaaagagaaacccgtatcccttaacctag 7015
|||||
Db 2250 TTGATGAATATGACTATTCGGTATTAACGACCAAAAGAAACCCGATCTTACCTTAG 2309
QY 7016 ctgtatgctcggataggaagccggttggcgtaggaacaggaagctgcccgtatca 7075
|||||
Db 2310 CTGTAAATGCTCGGATTAGGAGACGGCGTGGCTAGGACAGGACAGCTGCCCTGATCA 2369
QY 7076 caagacacacagcagctagagaagaaagctgtgtgagctataatcgcccatgacaagaatc 7135
|||||
Db 2370 CAGGACCAAGCAGCTAGAGAAAGACTGTGTGATCTATATGCTGACAGACAGAAATC 2429
QY 7136 tccgaagcctlaaagagagctgtttagcaacctagaagagccttgactcttgtctgaag 7195
|||||
Db 2430 TCGGACCTTTAGAGAGAGCTGTTAGCACTTGAAGAGCTTCTTGTGTGAG 2489
QY 7196 tggttctacagaacgggaagggatagatctgctgttcttaagagaaggtggtatgtg 7255
|||||
Db 2490 TGGTTCTACAGAACCGGAGGGATTAGATCTGTTCTTAAGAGAAGGTGGTATGTG 2549
QY 7256 cagccttaagaaagaaatgttctctatgtatgacatccatccaggaactcca 7315
|||||
Db 2550 CAGCTTTAAAGAGAAAGTGTCTTATGATGACTCTGAGACCCCTTCAAGACTCCA 2609
QY 7316 tgaacaagcttagaanaaagcttagagagcgctcgaaggaagaaagctgaacggggt 7375
|||||
Db 2610 TGAGCAAGCTTAGAGAAAGTTAGAGAGGCGTGAAGGAAAGAGAGCTGACAGGGGT 2669
QY 7376 ggtttgaagatgtgtcaacaggtccttgatgatagaacacccctgtctgtacag 7435
|||||
Db 2670 GGTTTGAAGAGATGTTCAACAGCTCTCTTGTATGACCAACCTGTCTCTGACGG 2729
QY 7436 ggccttagtagtctgtccctgttacttaagttgggcttgtcttaataagtttg 7495
|||||
Db 2730 GACCCCTAGTACTCTGCTCTCTGTTACTTACAGTTGGCCCTTGCTTAATTAATAGTTTG 2789
QY 7496 ttgcttgttagagaagagtgagtgcaagtcagataatgtaacttagaagaagtaacc 7555
|||||
Db 2790 TTGCTTTGTATGAGAAAGAGTAGAGTGCAGTCAGATCATGTGTAAGCAACAGTAC 2849
QY 7556 aagccttctgagcgaagaggaactgaaccttagccttcccaagttcaagataagact 7615
|||||
Db 2850 AAGGCTTCTGAGCCAAAGAGAACTGACCTTACCTCCCATCTTCAATTAATAGAACT 2909
QY 7616 attaaacagacagaagtgagggaatgaagatgaanaatgaacactaacctcccaaac 7675
|||||
Db 2910 ATTAAACAAGACAGAAGTGGGGAATGAAGATGAATAATGAACCTTACCTCCAGAAC 2969
QY 7676 ccaaggaagtaataaagactctaaatgcccccgaaattacagacccgtgtgctgcaat 7735
|||||
Db 2970 CCAGGAAGTTAATAAAGCTCTAAATGCCCGCAATTCAGAGACCTGTGTGCTGCCAGT 3029
QY 7736 aaatagtagaaggtcacacttctactatgttccagggcgtgtatcctgtgctaagtaag 7795
|||||

```
Db 3030 AATAGGTAGAGTGCACACTTCTATTGTTCCAGGGCCCTGCTATCCMGCCCTAAGTAAG 3089
QY 7796 ataacaggaatgagttgacctaatacgtatcgtgallctgtaaaactgactgacccat 7855
    |||||||
Db 3090 ATACAGGAATGATTGACTAATGCTTATCTGATTTCTGTAACCGACTGGCACCAT 3149
QY 7856 agaagaattgattacacattgacagccctagtgacctatctcaactgcaatctgcaactc 7915
    |||||||
Db 3150 AGAAGAATTGATTACACATTGACAGCCCTAGTACCTATCTCACTGCATCTGTCACTC 3209
QY 7916 tgcctcagagagccacagcaatgagagccctcggagctatlttaaatgattgltccacag 7975
    |||||||
Db 3210 TGCCCGAGAGCCACGAGATGCGGACCTCCGAGCTATTTAATGATTGTCACGCG 3269
QY 7976 agcgcgggctcgcgatatlttaaatgattgltccatgagagcgagctc-tcgatatlt 8034
    |||||||
Db 3270 AGCGCGGCTCTCGATATTTTAATGATTGTTGTCACAGGCGGCGCTCTCGATATTT 3329
QY 8035 taaatlgattglttgtaagcaagagcttltgtgaacccataaaagctgtccgat 8094
    |||||||
Db 3330 TAAATGATTGTTGTGACGACAGGCTTGTGTGAACCCCATTAAGCTGTCCCGAT 3389
QY 8095 tccgcactcgggagccgacatcctcctacccctcggtgltgtaagactgtgagccacagcg 8154
    |||||||
Db 3390 TCCGCACTCGGGCCCGAGTCTTAACCCCTGCTGTGTAAGACTGTGGGCCCCCAAGCG 3449
QY 8155 gcttgaataaaaaactcctctgtgtgtgcatc 8187
    |||||||
Db 3450 GCTTGGAAATAAAATCTCTTGTGCTTGTGCATC 3482
```

Search completed: February 24, 2002, 07:30:19
Job time: 28334 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:30:19 ; Search time 10173.5 Seconds

(without alignments)
4793.420 Million cell updates/sec

Title: US-09-171-553B-9

Percent score: 2956
Sequence: 1 tgccttttagggtagaac.....aaaaaaaaaaaaaaaaaa 2956

Scoring table:

OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:
1: gb_da:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_da:
16: em_fun:
17: em_hum:
18: em_in:
19: em_om:
20: em_or:
21: em_ov:
22: em_pat:
23: em_ph:
24: em_pl:
25: em_ro:
26: em_sts:
27: em_sy:
28: em_un:
29: em_vi:
30: em_htgo_hum:
31: em_htgo_inv:
32: em_htgo_rtd:
33: em_htg_hum:
34: em_htg_inv:
35: em_htg_rtd:
36: em_htg_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8
SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2191	74.1	8209	6	A66553	A66553 Sequence 3
2	2089	70.7	8196	6	A66552	A66552 Sequence 2
3	1872	63.3	1974	6	A66559	A66559 Sequence 9
4	1763	59.6	8750	14	PEN133818	AJ133818 Porcine e
5	1743	59.0	3482	6	AX002804	AX002804 Sequence
6	1743	59.0	3482	14	PERENY102	Y12239 Porcine end
7	1654	56.0	7808	14	PERY17013	Y17013 Porcine end
8	1636	55.3	8918	14	PEN133816	AJ133816 Porcine e
9	1572	53.2	8763	14	PEN293657	AJ293657 Porcine e
10	1499	50.7	1971	14	AF014162	AF014162 Porcine e
11	813	27.5	7333	4	AF038601	AF038601 Sus scrof
12	813	27.5	7333	6	AR130474	AR130474 Sequence
13	626	21.2	8849	14	PEN133817	AJ133817 Porcine e
14	457	15.5	2462	6	AX002802	AX002802 Sequence
15	457	15.5	2462	14	PERENY1	Y12238 Porcine end
16	446	15.1	1042	14	PERY17012	Y17012 Porcine end
17	412	13.9	2000	14	PEN288588	AJ288588 Porcine e
18	375	12.7	3025	4	AF147808	AF147808 Sus scrof
19	347	11.7	1968	14	PEN288589	AJ288589 Porcine e
20	321	10.9	746	14	PEN298073	AJ298073 Porcine e
21	318	10.8	590	14	PEN298074	AJ298074 Porcine e
22	318	10.8	2000	14	PEN288592	AJ288592 Porcine e
23	143	4.8	3320	6	A66551	A66551 Sequence 1
24	143	4.8	8918	14	PEN293656	AJ293656 Porcine e
25	139	4.7	2002	14	PEN288591	AJ288591 Porcine e
26	129	4.4	1996	14	PEN288590	AJ288590 Porcine e
27	116	3.9	1999	14	PEN288587	AJ288587 Porcine e
28	98	3.3	8132	4	AF038600	AF038600 Sus scrof
29	98	3.3	8132	6	AR130475	AR130475 Sequence
30	91	3.1	633	6	AX052622	AX052622 Sequence
31	91	3.1	704	6	AX052621	AX052621 Sequence
32	91	3.1	4918	6	AX052637	AX052637 Sequence
33	91	3.1	6076	6	AX052636	AX052636 Sequence
34	91	3.1	6076	6	AX052647	AX052647 Sequence
35	91	3.1	7362	6	AX052634	AX052634 Sequence
36	91	3.1	7873	6	AX052638	AX052638 Sequence
37	91	3.1	8060	6	AR130473	AR130473 Sequence
38	91	3.1	8132	4	AF038599	AF038599 Sus scrof
39	88	3.0	2008	14	PEN288586	AJ288586 Porcine e
40	74	2.5	4402	6	AX052635	AX052635 Sequence
41	62	2.1	1996	14	PEN288584	AJ288584 Porcine e
42	57	1.9	610	4	AF286463	AF286463 Sus scrof
43	53	1.8	595	14	AF298075	AJ298075 Porcine e
44	41	1.4	1591	10	AF161307	AF161307 Mus carol
45	41	1.4	195423	2	AC087151	AC087151 Mus muscu

ALIGNMENTS

RESULT 1
A66553 8209 bp DNA PAT 29-MAR-1999
LOCUS Sequence 3 from Patent WO9740167.
DEFINITION A66553
ACCESSION A66553.1 GI:4538106
VERSION
KEYWORDS
SOURCE
ORGANISM

porcine endogenous retrovirus.
porcine endogenous retrovirus
Viruses; Retroid viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.

REFERENCE
1 (bases 1 to 8209)
AUTHORS Galbraith,D.N., Haworth,C., Lees,G.M. and Smith,K.T.
TITLE PORCINE RETROVIRUS
JOURNAL Patent: WO 9740167-A 3 30-OCT-1997;
Q ONE BIOTECH LTD (GB)

FEATURES
source
1..8209
Location/Qualifiers
/organism="porcine endogenous retrovirus"
/db_xref="taxon:61673"

CS 588..2162
/codon_start=1
/product="VIRAL CORE PROTEIN"
/protein_id="CAB39347.1"
/db_xref="GI:4538107"
/translation="MGQVITPLSLTLDHWTVEVRNANLSVQKKGMPQWQFCASEMP
TDPVMPSEGTFSNELLAVKAIIFQOTGSHDDEPVLITWQDAEDPPVAVKPMN
KPKPGPRLILALGKKNHSAKEVSESSVLPDRGADLAGTPCSPPLSTGCCG
TSAPGAPVVEGPAQTRSRGATPERDEIALLPLRTYMPGGLQLOPYWPS
ADLYNKTTHNPPSEDPORLTGLVSMSSHOPMDQCOLLOTLPTEERRILLRA
RKVPGADRPTOLONEIDMGFPLTRPGDYTAAGREBKITYRQALVAGLSAARRP
TNLAKREVWQGNPEPVSFLRMEAFRRFPDPTSAQASVALATGSDAIDIR
KTLAAVEGKSSRRERDRKIRSGRQSGNLGNTPLDKQCAVCKEKGHWARNCPK
KGNKPKVLALEDKND"
<2163..5747
/codon_start=1
/product="REVERSE TRANSCRIPTASE"
/protein_id="CAB39348.1"
/db_xref="GI:4538108"
/translation="GRGSDPLPEPRVTLKVEGQPEFLVDGAEHSLLOPLGKIKE
KSMWGATGQROYPWTRRTVDIGVGRTHSEIVIPCEPVALRDILTKKGAISTF
BOGREPVSNKKRITYITLQIDDEVRLSPYQKPODIDSLMEQFQANAEFAGMGLA
KQVPOVIOLKASATPVSVRQYPLSREAREGLMPHVRILQGLLVQSPWNTPLP
VRKPGINDYRPVQDLREVNKRVODIHPVNPYNLISALPPERMVTVLIDKAEFCL
RLHPTSOPLFAFEMWRDPTGRTGOLTWRLPQGFNSPTIFEDALHRLDIANRQHPQ
VTLLQYVDDLLLAGATKODLEGTKALLELSDLGABAKAJOICRREVTLAGSLR
GGORMLTEARKKTYVOIPAPTAQKQVRELTGATGCRILMIPGATLALVPLTEBK
GSMAPHEOKADAIKALLSAPALAPVPTKFLYDERKGVAVITQTLGMR
PVAUSKRLDPVAGMPVCLKATAVAAILVKADKLLTQNTITVAPHALENTVQPP
DRMTNARMTYQSLTLTERTVFAPPAALPATLLPEETDEVTQDCHQLIEETGR
KDLTIDPLTGEVLTWFTDSSVVEGKRNAGAAYVDGTETIASSLPBSTSKQKELM
ALTOALRLAEKGSINITYDSRYAFATAYHAGALYKQGLTISAGREIKKEILSLE
ALHPLKRLAIHCPGHOKAKDLISRGNOADHVAQAOAVNLPLIEPKAPPEPRO
YTLDEMOELKIDQSEPEEGTCYSYGEITLPHKEGLEYOHRHTLPTGKHQOL
VRTSPYHVLRLPEVADSVYKPCPOLVYANPSRIPGKRLGSHGAMWEDPTVEK
PAKYGNKYLVFVDFTSVVEAVPTKERTSTVAKLIEELIRFPGIPEISDSQGA
FVAQVSOGLAKILIGIDMKLHCAYRPOSSQVERMNTIETLLKLTTEGKLEMMALL
PEVLFVRNTPGOGLPYELLYGPPALFAISAVLISOPLESGLKLEVRQ
RAWKQIREAGSDLOVPHRFQVDSVYRRRAGLLEFRMKGPYVILLTPTAVAKE
GIPTWTHASHVKKPAPPDSDGMAEKENTLKLRLHRLVYSNNNSQO"
5620..7590
/codon_start=1
/product="VIRAL ENVELOPE PROTEIN"
/protein_id="CAB39349.1"
/db_xref="GI:4538109"
/translation="MHPILSRHLPTRGGEPRKRLRPLSFASIAMFLTLTPQASK
RLIDSNPHRPLSLTMLTIDPNTGVNSTRGVAAPGCTWMPHLCRLINPAVSTP
PNLVSTGYCCDGTREKCYCGSGSEFRRNSCVTSNDGKMPISLIDRYKESFVN
SGGKTKMMKLYKDKSCSPSDLYLKISTEKTKGYSKVDWYELGNSFLLYGAGAG
SLLTILRIETGTEPPVAMPKPVLAEOGPALPEPHNLVPQLTSLRDITQPPNS
TTLTPTNTPRNSGPVAKTGORFLSIGAFOALINSTDPATSSCMLTSSGPPYE
GMAKERFVYTKHNOCTMGSRNKLTLTEVSGKICIKAPPSHCLTSVYVEOA
SENOYLVPGYNMMAACNGLTPCVSYVSNOKRDVMAIPIRYVYHEEVLEVDY
YRNRKRPVSLTTLAVMLGLTAVGVTGCTAALLTPPOULEKGEJLHAANTEDIRA
LKEVSINLESLSLSSEVVLQNRGLDLFLREGGCAALKKECCGYVDHSAIDISM
NKLRLKERRRERREADQGEFENRSPWMTLLSALGPLVILLTLTVPCLINR
FVAFVERYSAVQINVLROOYGLISQSETDL"

BASE COUNT 2168 a 2064 c 2154 g 1823 t
ORIGIN

Query Match 74.1%; Score 2191; DB 6; Length 8209;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2741; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 208 gagactgcaagtcacacogcttccaagtgtgagattcaagtctatgttagagccac 267
DB 5464 GAGACTTGCAAGTTCACACATCGCTTCCAAAGTTGAGATTCAGTCTATGAGACGCCACC 5523
QY 268 gtgcaggaacacgcagacccgtgtggaaggacattaccogtactttgacacacaa 327
DB 5524 GTGCAGGAACCTCGAAGCTCGTGTGGAAGGACCTTATCTCGTACTTTTACACACACCA 5583

QY 328 cggctgtgaagtcgaaggaatctccacctggtatccatgcatcccaagtltaagctgagc 387
DB 5584 CGGCTGTGAAGTTCGAGGAATCCCCACCTGGATCCATGATCCACAGCTTAAGCGCGGC 5643
QY 388 caactccgactcggggtgtgagagccgaagaagacttgaaatcccttaagcttgcctcc 447
DB 5644 CACCTCCGACACTCGGGGTGGAGAGCGGAAGACGAAATCCCTTAAAGCTTGGCCCTCC 5703
QY 448 atcgctggttcttaacttaacaaataactcccgagccagtaagtaacgcttaagac 507
DB 5704 ATCGCTGAGTTCCTTACTTAACTTAACATACTCCCGAGCCAGTAGTAACGCCCTTATGAC 5763
QY 508 agctgaacccccatagacacttatacccttaacctgctgattatgacccctgatacgggt 567
DB 5764 ACTCGAACCCTCATAGACCTTATCCCTTACCTGGTGGATATGACCTGATACCGGCT 5823
QY 568 gtcaactgtaataagcaactcgaagtgctgtcctctagaagcaactggtgagcgaactgat 627
DB 5824 GTCACTGTAAATAGCACTCGAGGTGTCTCTCTAGAGGACCTGTGTGCTGAACTGCAT 5883
QY 628 tctctcctccgattgattaaaccgctgtttaaagcacacctcccaactagtcctgagt 687
DB 5884 TTCTGCTCTCGATTGATTAAACCCCTGTAAAGCACACTCCCAACTGATCCGTAGT 5943
QY 688 tatgggttctatctgtcccaaggcaagaagaagaatactgtg99ggtcttgg99aa 747
DB 5944 TATGGGTTCTATGCTGCCACAGCACAGAAAGAAATGACTGTGGGGTTCTGGGAA 6003
QY 748 tccctctgtaggagatgagactgctgctcaactccacgcttgagacttggaatgcccga 807
DB 6004 TCCCTCTGTAGGAGATGAGCTGCTGCTCACTCCACGATGAGACTGGAAATGCGCGATC 6063
QY 808 tctctccagacgggtataaatctcctcttgcaatctccgcccgggaagtacaaatg 867
DB 6064 TCTCTCCAGGACCGGGTAAATTTCTCTTTGCAATTCGGCGCCGGCAAGTACAAATG 6123
QY 868 atgaaactatataagataagagctgctcccccatacgaacttagattatctaagaatag 927
DB 6124 ATGAAACTATATAAGATTAAGAGCTGCTCCCATCAGACTTAAGATTATTAAGATAGT 6183
QY 928 ttcactgaaagaaagaaacagaaataatctaaagtgattaaatgattagctgaggga 987
DB 6184 TTCACCTG--AAAGGAAAGAGAAATATTTCAAAGTGGATTAAGTATGATGAGCTGGGA 6241
QY 988 atagttttttaaataatgctgaggggagcaggtgctcaactttaaaccatcgcttagata 1047
DB 6242 ATAGTTTTTT-ATTATATGCGGGGAGAGAGGTCACATTAAACATTGCGCTTAGATA 6300
QY 1048 gagaegggagacgaacccctgtgtgcaatgggaacccgataaagtaactgtgctgaaacagg 1107
DB 6301 GAGACGGGACAGAACCCCTGTGCAATGGACCCGATAAAGTACTGGCTGAACAGGGG 6360
QY 1108 ccccgagcccttgaagacacacgaataacttgcggtgccccaaatlaactgcttggagct 1167
DB 6361 CCCCAGCCCTGAGACACCCGATTAACCTTGCGGATGGACCCGATAAAGTACTGGCTGAACAGGGG 6420
QY 1168 gacataaacacagcgccttagacaaagtaacacatgtagattctactacaacagcctaga 1227
DB 6421 GACATTAACACAGCGCGCTGACAAAGTACCACTGATTTGATTCCTAACCAACAGCCTAGA 6480
QY 1228 aactcccaagtgcttctgtttaaagacaggaagaacacttcaagtctcaacagggagct 1287
DB 6481 AACTCCCAAGTGTCTCTGTTAAAGACAGACACTTCAAGTCACTATCCAGGAGACT 6540
QY 1288 ttcgaagcataaactcaacacagccttagacacacacacacacacacacacacacacacac 1347
DB 6541 TTCGAGGCTATTAACCTCAACCACTGATGCACTTCTTCTGTTGGCTTTGTATACC 6600
QY 1348 tcaaggctcccttatatagagggaatgagctaaagaagaanaatltcaatgtgaccaaagag 1407
DB 6601 TCAGGCGCTCTTATATATAGAGGAGATGCTTAAGAAAGAAATTCATATGTGACCAAGAG 6660

QY 1408 catagaatcaatgacatgaggtgcccgaatlaagcttaacctcaactgaagttccggg 1467
 |||||
 Db 6661 CATAGAAATCATGTATGACATGGGGGTCCGAAATAGCTTACCTCAGTGAAGTTCCGGG 6720
 |||||
 QY 1468 aaggagaaatgacatgagaaagctcccccaccccaaacacttgctataagcttg 1527
 |||||
 Db 6721 AAGGGACATGACATAGGAAAGCTCCCATCCACCAACACCTTGTATAGTACTGG 6780
 |||||
 QY 1528 gtttaagcagagccctcagaanaatcagtaattagtaacctggtataacagtgctgggca 1587
 |||||
 Db 6781 GTTTATGACAGCCCTCACAATAATCATGTATTAGTACCTGTTATTAACAGTGGTGGCA 6840
 |||||
 QY 1588 tgcatactggtgtaacccccctggttccacctgaagttccaaccaatccaagatttg 1647
 |||||
 Db 6841 TGCATATGAGGTTTACCCCTGTGTTCACCTCAGTCTTCACCAATCCAAAGATTTTC 6900
 |||||
 QY 1648 tggtaagtgcccaatgctcccccagtgtaacctacatccctgaggaagtgctctgat 1707
 |||||
 Db 6901 TGTGTATGCTCAATTCGTCCTCCGAGTGTACTACCATCTGAGAGATGGTCTTGTAT 6960
 |||||
 QY 1708 gaatagaactatcgatataaccgacaaagaagaaacccgtalcccttaacctagctgta 1767
 |||||
 Db 6961 GAATATGACATATGGTATATACCGACCAAAAGAGAACCCGTATCCCTTACCTAGCTGTA 7020
 |||||
 QY 1768 atgctcggatlaaggaagcgcgttggtgagaaacggagacgctgctctgatacagga 1827
 |||||
 Db 7021 ATGCTCGATTTAGGAGCGGCGCTTGCGTGAAGAACGAGACACTCCCTGATCACAGA 7080
 |||||
 QY 1828 cccagagagctagaagaagacttggtgagctacatgagcgccatgacagaagaatccgca 1887
 |||||
 Db 7081 CCACAGCACTAGAGAAAGAGACTTGGTGTACCTATGCGGCCCATGACAGAGATCTCCGA 7140
 |||||
 QY 1888 gccctaaaggagctggttagcaacctagaagagtcctgactcttctgtctgaagtggt 1947
 |||||
 Db 7141 GCCTTAAAGAGAGTCTTTAGCAACCTAGAGAGTCCCTGTCTTGTCTGAAGTGGT 7200
 |||||
 QY 1948 ctacagaacgggaggtatagatctgctgttcttaagaagaagtggtatgtgagcc 2007
 |||||
 Db 7201 CTACAGAACCGAGGAGGATGATCTGCTGTTCTAAGAAAGTGGGTTATGTGACGCC 7260
 |||||
 QY 2008 ttaaaaaagaatgttgctctatgtagatcaactcagagagccatcagagactccatgaac 2067
 |||||
 Db 7261 TTTAAAGAGAGATGTCTCTTATGTATGATCACTCAGAGCCATCAGAGACTCATGATAC 7320
 |||||
 QY 2068 aagcttagaaaaagttagagagcgctcgaaggaaggaagagctgcacggggtgtt 2127
 |||||
 Db 7321 AAGCTTAGAAAAAGTTAGAGAGCGCTCGAAGGAGAAAGAGCGCTGACCGGGGTGTT 7380
 |||||
 QY 2128 gaagagatggttcaacaggtctccttgatgacacacccctgcttctgctgacggggcc 2187
 |||||
 Db 7381 GAAGGATGCTCAACAGGTCTCTTGTGATGACACCTGCTTCTGCTGACGGGGCC 7440
 |||||
 QY 2188 ctatagctcgtcctgcttgtaactacagttgagccttgcttaataagttgtgtgc 2247
 |||||
 Db 7441 CTAGTAGCTCTGCTCTCTTACTTACAGTTGGGCTTGTCTTAATTAATAGGTTGTGTGC 7500
 |||||
 QY 2248 ttgttagagaagagtgatgagcgtccagatcagtgtagtactagaagaagttacaaagc 2307
 |||||
 Db 7501 TTTGTTAGAGAACGAGTAGAGTGCAGTCCAGATCATGTAGGACACACATACCAAGGC 7560
 |||||
 QY 2308 ctctgagcagaaggaagac 2367
 |||||
 Db 7561 CTTTGAGACCAAGAGGAAACCTGACCTTACCTCCAGTTCTTAAGATTAGACATATTA 7620
 |||||
 QY 2368 caagaagaagtggtggaatgaagaatgaagaatgaagaatgaagaatgaagaatgaaga 2427
 |||||
 Db 7621 CAAAGCAAGAGTGGGAGATGAAGAGATGAAGAGATGAAGAGATGAAGAGATGAAGAG 7680
 |||||
 QY 2428 aagtaataaagaagcttaaatgcccccaattacagaacccctgctgctgacgttaata 2487
 |||||
 Db 7681 AAGTTAATAAAGAGCTCTTAATGCCCCCGAATTAAGACCTGCTGCTGCTGCTGCTGCT 7740
 |||||
 QY 2488 ggtagaaggtcacacttactatgttccagggcgtgctatccctgagtaagtaagtaac 2547

Db 7741 GGTGAAAGGTACACCTTCTTATTTTCCAGGGCTGCTATTCGCGCTAGATGAATTAAC 7800
 |||||
 QY 2548 aggaatgagtgtaactatcgtctatctgagcttctgtaaaacgtaactggaacaaagaag 2607
 |||||
 Db 7801 AGGAATAGCTTGAATATGACCTTATCTGATGCTGTAATACTGACGACCACTTAAG 7860
 |||||
 QY 2608 aattgataacacattgacagccctgagtgacatctacacacacacacacacacacacac 2667
 |||||
 Db 7861 AATTGATTACACATTTGACAGCCCTAGTACCTATCTCAACATGCAATCTGTACCTGCC 7920
 |||||
 QY 2668 aggaacccacgacagatgagacacacacacacacacacacacacacacacacacacac 2727
 |||||
 Db 7921 AGGACCCACGACATGCGAGCTCCGAGCTATTTTAAATGATGGTGTCCACGACGCG 7980
 |||||
 QY 2728 gggctctgataatttaaaatgattgctgacatgagcgggacccatgataattaaat 2787
 |||||
 Db 7981 GGGCTCTGATATTTTAAATGATGGTGTCCATGAGCGCGGCTCTGATATTTTAAAT 8040
 |||||
 QY 2788 gatgagtgtagacacagagcttggtgtaaccccaataaagctgtccgattccgca 2847
 |||||
 Db 8041 GATTGCTTGTGACCCACAGCGCTTGTGTGAACCCCATTAAGCTGTCCGATTCGCA 8100
 |||||
 QY 2848 ctgagggcgcagctcctcctacccctggtggtgtaacagctggtggtggtggtggtggt 2907
 |||||
 Db 8101 CTGGGGCGCGAGCTCTTACCCCTGCGTGTGAGACTGTGGCGCCACGCGCTTGG 8160
 |||||
 QY 2908 aataaaatccctctgctggttgatcaaaaaaataaaaaaataaaaaa 2956
 |||||
 Db 8161 AATTAATAATCTCTGCTGTGATCAAAAAAATAAAAAAATAAAAAA 8209
 |||||

RESULT 2
 A66552 8196 bp DNA PAT 29-MAR-1999
 LOCUS A66552
 DEFINITION Sequence 2 from Patent WO9740167.
 ACCESSION A66552
 VERSION A66552.1 GI:4538105
 KEYWORDS
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 porcine endogenous retrovirus
 Viruses; Retroid viruses; Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 REFERENCE
 1 (bases 1 to 8196)
 Galbraith, D.N., Haworth, C., Lees, G.M. and Smith, K.T.
 PORCINE RETROVIRUS
 Patent: WO 9740167-A 2 30-OCT-1997;
 JOURNAL Q ONE BIOTECH LTD (GB)
 FEATURES
 source
 1..8196
 /organism="porcine endogenous retrovirus"
 /db_xref="taxon:61673"
 BASE COUNT 2165 a 2061 c 2146 g 1820 t 4 others
 ORIGIN

Query Match 70.7%; Score 2089; DB 6; Length 8196;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 2739; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

QY 208 gaaactgcaagttccacatcgcttccaaattgagatcagctatgtaagcagacc 267
 |||||
 Db 5450 GAGACTGTGCAAGTTCACATCGCTTCCAAAGTTGAGATTCATGTATGAGCGCAC 5509
 |||||
 QY 268 gtgcagaaacccctcgaagctcgglygaagagacacttaactcgtactttgacacacaa 327
 |||||
 Db 5510 GTGAGAGAAACCTCGAATCGGAGGAGGAGACTTATCTGTAATTTGACACACAA 5569
 |||||
 QY 328 cggctgtgaaagtcgaagaaatctccacccgagatcatgataccacgttaagctggcg 387
 |||||
 Db 5570 CGGCTTGAAAGTGAAGGAATCCCCACCTGAGATCCATGCACTCCACGTTAAAGCTGGCG 5629
 |||||
 QY 388 caactccgcactcgggttgagagacgaaagactgagaatcccttaagcttcgctcc 447

Db 5630 CACCTCCGACTCGGGGGAGAGCCGAAAGACTGAGAAATCCCTTAAGCTTCCGCTCC 5689
QY 448 atcgcttgcttactactactaactaactccccagccagtaataacgcttatagac 507
Db 5690 ATGGCTGGTTCCTTACTACTACATAATCTCCAGGCGAGTATTAACGGCTTATAGAC 5749
QY 508 agctcgaaccccaatagacttlaacccactctgctgtatattgaacccatagacgt 567
Db 5750 AGCTCGAAGCCCGCATFAGACCTTTATCCCTTAACCTGCTGATTAATGACCTGATACGGCT 5809
QY 568 gtaacttlaaataagactgaaggtgtgcttctctagagacccggtggtcgtgaatcat 627
Db 5810 GTCACTGTAATTAACACTCGAGGTGTGCTCTAGAGGACACGTGGCTCGAATGCTCAT 5869
QY 628 tctgctccgattgattaaaccccgctgtttaaagacacactcccaactgactgact 687
Db 5870 TTTGCTCCGATGATTAAACCCCGCTGTAAATAGCACACCTCCCAACCTAGTCCGTACT 5929
QY 688 tatgggtctattgtctgcccagacacagaaagaaataactgtgggttctgaggaa 747
Db 5930 TATGGGTTCTATTGCTGCTCCAGGACACAGAAAGAAATACGTGGGGTCTCGGGAA 5989
QY 748 tctctctgtagagatgagctgtctcaactccaacgattgagactggaatgtccgact 807
Db 5990 TCCCTTGTAGAGATGAGACTCGTCACTCCACGATGAGACTGGAATGCGGATC 6049
QY 808 tctctcagagacccggttaaaattctccttctgtaattccggtccggccaagtacaaatg 867
Db 6050 TCTCTCCAGACCGGGTAAATATCTCTTGTCAATTCGGCGCGGCAAGTACAAATG 6109
QY 868 atgaactataaagataagagctgctcccaactcagacttgattatataagataagt 927
Db 6110 ATCAACTATATATAATAGAGACTCTCCCACTCAGACTGTAGATTATTTAAGATAAGT 6169
QY 928 ttaactgaaaaaagaaacagaaataatcaaaagtgtgataatgataatgagtgagga 987
Db 6170 TTCTACTG- AAAGGAAAGAGAAATATTCAAAAGTGTATTAATGATGAGCTGGGGA 6227
QY 988 atagtttttaataataatgaggggagagaggtccacttaacattgccttagata 1047
Db 6228 ATATGTTTTT- ATATATATGCGGGGAGAGAGGCTCCACTTAAACATTGCGCTTAGATA 6286
QY 1048 gagaacgggagacagaaacccctgtgcaatgtgacccgataaagtaactgtgtgaacagg 1107
Db 6287 GAGAGCGGAGACAGAAACCCCTGTGGCAATGAGAACCCGATTAAGTACTGCTGAACAGGG 6346
QY 1108 ccccggtccctgtgagccacgcataactgtccgtgtcccaattlaacctgctgagct 1167
Db 6347 CCCCCGGCCCTGAGCCACCCGATTAATGCGGTGCGCCCAATTAACTCGCTGGGCT 6406
QY 1168 gacataacacagccctagacaaagtaacacttgattgatttcttaaccaacagcctaga 1227
Db 6407 GACATTAACAGACCCCTCTAGCAACAGTACCACTGATTGATTCCTAACCAACAGCCTTGA 6466
QY 1228 aactcccaagtgctcctgttaagacagagacagactctcaqlctcatccagagact 1287
Db 6467 AACTCCCAAGGTCTCTGTAAAGACAGACAGAGACTCTCAAGTCACTATCCAGAGACT 6526
QY 1288 ttccaagcattcaactcaacgacactgtatgccaacttctgtgtgctgtctatcc 1347
Db 6527 TTCCAAGCATTCACTCAACGACCCCTGATGCTCACTCTTCTGTGCTTTGCTTATCC 6586
QY 1348 tcaaggctccttattatgagggagatgcttaagaagaataattcaattgtgacaaagag 1407
Db 6587 TCAGGGCTCTTATTATAGAGGAGATGCTTAAGAAAGAAATTAATGATGACCAAGAG 6646
QY 1408 catagaaatcaatgataaggggttccgaataagcttaacccatctgaatgtccgg 1467
Db 6647 CATGAATCAATGATACATGGGGTCCGGAATTAAGCTTACCTCATGAAATTTCCGGG 6706
QY 1468 aaggggcatgcatagagaaagctccccatcccaacacacttgcataagactgtg 1527
Db 6707 AAGGGATATGATAGAAAAAGCTCCCATCCCAACACCTTTGATATGATCTGTG 6766
QY 1528 gttatgagcagggcctcagaanaatcagiatltaglaactcgtgtataacaggtgtggga 1587
Db 6767 GTTATAGAGAGGCTCAGAAATCAGTATTTAGTACCTGGTATTATACAGGTGTGGGA 6826
QY 1588 tgcataactgggtlaacccccgtgttctcacactagttctcaaccaatccaagattg 1647
Db 6827 TGCATATAGGGGTAAACCCCTGTGTTCACCTCACTGCTTCAACCAATCCAAATATTC 6886
QY 1648 tgttcatgtccaaatcgtcccccaggtgtactaccatccctgaggaaggttctctat 1707
Db 6887 TGTGTCATGCTCAATATGCTCCCGAGTGTACTACATCTGAGAGAGTGTCTTAT 6946
QY 1708 gaatatgactatcgtatataacgacccaaagagaccgtlaacccatccatagctgta 1767
Db 6947 GAATATAGCTATTCGATTAACCGACCAAAAGAAACCGTATCCCTTAACCTACTGTA 7006
QY 1768 atgctcgaattaggaagcggttggtgtaggaagagagagagagagagagagagagag 1827
Db 7007 ATGCTCGATTTAGGACGGCGCTTGGCGTAGAACAGGACAGCTGCCCTGATACAGGA 7066
QY 1828 ccacagcagctagagaaagacttggtagctatactgcgcataagacagatctccga 1887
Db 7067 CCACAGCAGCTAGAGAAAGACTTGTGAGCTACATCGGCGCATGACAGAAATCTCGA 7126
QY 1888 gcttaagaggtctgttagaacacttaagaagctcctgacttctgtcgaatgtgt 1947
Db 7127 GCCTTAAAGAGATCTGTAGAACCTTAAGAGTCCGACTTCTTGTCTGAAGTGTGT 7186
QY 1948 ctacagacccgagagggatagatctgttcttaagaaggttgggttctgagggc 2007
Db 7187 CTACGAAACCGAGGGGATTTAGATCTGCTTTCTTAAGAGAGTGGGTATGTGCAAGC 7246
QY 2008 ttaaaagaagaatgttcttctatgtagataactcagagagacacagactccatgaac 2067
Db 7247 TTAAGAGAGATGTGCTTATATGATGATCACTCAGAGACCATGAGACTCATGAC 7306
QY 2068 aagcttaaaaaagttagaagaagcgtcgaaggaaagagagaggtgcccaggggttct 2127
Db 7307 AAGCTTAAGAAAAAGTTTATGAGAGGCTGCAAGGGAAGAGAGGCTGACAGGGGTGCTT 7366
QY 2128 gaagatgtgtcaacaggtctccttgaatgaccccttcttctgtcgaagggccc 2187
Db 7367 GAAAGATGTTCAACAGATCTCTTGTGATGACCACTGCTTCTGCTGAGAGGGGCC 7426
QY 2188 ctatagctcgtcctcgttactactagctgtggcctgtcttaataaggtgtgccc 2247
Db 7427 CTATAGTCTGCTCTGTTACTTACAGTTGGGCTTGTAAATTAATAGTTGTGGCC 7486
QY 2248 ttgttgaagaacagtgatgacgtcagatcatatgttacttgaagaagagagagagc 2307
Db 7487 TTTGTATAGAGAGAGTGAATGCACTGATGATGATGATGATGATGATGATGATGATGAT 7546
QY 2308 ctcttgagccaagagagaactgacctcctagccttccaggttcaagaatagaactttaa 2367
Db 7547 CTCTGAGCCAGAGAGAACTGACCTTACCTTCCAGTTGATTAAGATTAAGACTATTA 7606
QY 2368 caagacaagaagtggaagaaatgaagaatgaagaatgaagaatgaagaatgaagaatga 2427
Db 7607 CAAGACAGAGAGTGGGGAATGAAGAGATGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 7666
QY 2428 aagtataaaaaagcttlaaattgccccgaattacagacccgtgtgtcgcagtaataa 2487
Db 7667 AAGTTATATAAAGCTCTAATATCCCGGATTTMAGACCTGCTGCTGCTGCTGCTGCTGCT 7726
QY 2488 ggtagaagttcaacttctatgttccagggcgtgtacttctgtgtccttaagaataaac 2547
Db 7727 GGTAGAAGTCAACACTCTATTTGTTCCAGGGCTGCTATCTGGGCTTAAGTAAGTAAAC 7786
QY 2548 aggaataggttgaactacgttctatctgtatctgttaaacactgtgacacatagaag 2607
Db 7787 AGGAATGAGTTGACTTAATGCTTATCTGATTTGTATAAACTGACTGGCACCATTAAGAG 7846

Qy	2608	aaagatataaacattgacagcccttaatgacatactcaactgcaatctgtaactctgccc	2667
Db	7847	AATTGATTACACATGTAGACACCCCTAGAGACCTATCTCAACTGCATCTGTCACTGTGCC	7906
Qy	2668	aggaagccacagcagatcgcgacctccgagctactttaaattgatatgltccagagcgc	2727
Db	7907	AGGAGCCACAGCANATCGGAGCCTCGGAGACTATTTTAAATGATTGGTCCAGGAGGC	7966
Qy	2728	gggctctcgatattttaaattgatatgltccatgagcgcgggctctcgatatltttaat	2787
Db	7967	GGGCTCTCGATATTTTAAATGATTGGTCCATGAGACGGCGGCTCTCGATATTTTAAAT	8026
Qy	2788	gattggttttggagaccagcgtttgttggaaaccccataaagaactgtccgattccga	2847
Db	8027	GATTGTTTGGAGCCACGAGCCTTTGTTGGAAACCCCATTAAGCTGTCCGATTCCGA	8086
Qy	2848	ctcgagcgcgagctccctactccctcgltglttaacgactgtggcccgagcgcgttgg	2907
Db	8087	CTCGGGGCCGAGTCTCTACCCCTCGTGtGTACGACTGTGGGCCCGAGCGCGTTGG	8146
Qy	2908	aataaaatccctctgtctgttgcatacaaaaaaataaaaaaataaaaaa	2956
Db	8147	AATTAATTCCTCTTGTGCTGTTTGCATCAACAAAAAATAAAAAAATAAAAAA	8195
RESULT	3		
A66559			
LOCUS	A66559	1974 bp	DNA
DEFINITION		Sequence 9 from Patent WO9740167.	
ACCESSION	A66559		
VERSION	A66559.1	GI:4538112	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT			
ORIGIN			

Query Match	63.33	Score 1872	DB 6	Length 1974
Best Local Similarity	99.99	Prod. No. 0		
Matches 1972: Conservative	0	Mismatches 2	Indels 0	Gaps
Qy 364 atgcattcccaagtttaagctgycgcacatcccgactcggggtgagagaccgaaagactg	423			
Db 1 atgcattcccaagtttaagctgycgcacatcccgactcggggtgagagaccgaaagactg	60			
Qy 424 agaattcccttaagctcgcgcacatcgcgcgtggttcctacttaacaataactcccgag	483			
Db 61 agaattcccttaagctcgcgcacatcgcgcgtggttcctacttaacaataactcccgag	120			
Qy 484 gccagtagtaaacgccttatagacagctcgaaeccecatagacattatccctactcgg	543			
Db 121 gccagtagtaaacgccttatagacagctcgaaeccecatagacattatccctactcgg	180			
Qy 544 ctgattatcgaccctgatacgaagctgctacatgtaataagcaactcgaggtgtgtccctaga	603			
Db 181 ctgattatgtaacccgatacgaagctgctacatgtaataagcaactcgaggtgtgtccctaga	240			
Qy 604 ggcaccctgtgtgctcgaactcattctcgtcctcgatlgatttaaccccgctgttaanaag	663			
Db 241 ggcaccctgtgtgctcgaactcattctcgtcctcgatlgatttaaccccgctgttaanaag	300			

QY	664	aaccttcacaaactbactccgtaagttatgggtttctatctgctcccgagcaagaaagag	723
Db	301	ACACCTCCCAACACTGACCTCGTAATTATGGGTTCTATTGCTGCCACGACAGAAAGAG	360
QY	724	aaatactgtaggggtctctggggaatccctctctgtagagatgtagagctgctcactccaac	783
Db	361	AAATACCTGTGGGGTTCTGTGGGAATCTCTGTAGAGAGATGAGCTGTGCTACCTCCAAC	420
QY	784	gatggagaactggaatgagccgacatctctcccaagaccggtataaatctctcttgtcaat	843
Db	421	GATGGAGACTGGAAATGGCCGATCTCTCCAGAGACCGGTTAAATTCCTTTGTCAAT	480
QY	844	tcgcgccccgggcaagfatacaaaatgatggaacatctataagataaagagtgctcccatca	903
Db	481	TCCGCCCCGGGCAAGTACAAATGATGAACATATATAAAGATMAAGCTGCTCCCATCA	540
QY	904	gacttagattatctaaagaataagtttccacttgaaaaaagaaaaacagaaatatctcaaaag	963
Db	541	GACTTTAACTTATCTTAAAGATTAAGTTTCACTGAAAAAGCAAAACAGAAATATTCCAAAG	600
QY	964	tgaataaattglatgagcttggggaaatagtttttlaataatgycgggggagcgaggtcc	1023
Db	601	TGGAFTAAATGGTATGAGCTGGGGAATAGTTTTTTATTAATATAGCGGGGAGAGGGTCC	660
QY	1024	aacttaaccattgcgcttaagatagagacggggagcaagaacccctgtgcaataggagacc	1083
Db	661	ACTTTAAACCAATTCGCTTTAGAGATAGAGCGGGACAGAAACCCCTGTGGCAATGGGACC	720
QY	1084	gataaagtactggttgaacagggggcccgcccttgagacacgcgtataacttgcggtg	1143
Db	721	GATTAAGTACTGGCTGTACAGGGGGCCCCCGGCTTGAGACCACCCGATTAAGTTGCCGTG	780
QY	1144	ccccaattaaacctgcgcgcgccttgacataaacaacagccgcctlagacaacagtaaccatgga	1203
Db	781	CCCCAATTAACTCGCTGGCGGCTGTACATACACAGCCGCTAGCAACAGTACCACTGGA	840
QY	1204	ttgattccaccaacacgcgcctagaacatccccaagtggtctctgttaagaacaggagacaga	1263
Db	841	TTGATTCCTTACCAACACGCTTAGAATCTCCCGAGGTGTCCTGTTAAAGACAGAGAGGA	900
QY	1264	ctcttaagtcctatccagggagagctcttccaaagcaataactaccagccctgaatgcact	1323
Db	901	CTCTTCAGTCTCATCCAGGAGGCTTTCCAGGCATTAATCTCCACCGACCTGATGCCACT	960
QY	1324	tctctctgttgcttctatcctccaagggccctctatctatgaggggagatggtctaaagaa	1383
Db	961	TCTTCTGTGTGGCTTTGTCTCATCTCCAGGGCCCTCTATTATTAGGGGATGGCTAAAGAA	1020
QY	1384	agaaaattcaatgtgaccaaaagacabagaatcaaalgtacatbggggtcccgaaataag	1443
Db	1021	AGAAATTCATATGTGACCAAAAGACATATGAAATCATGTATGAGGGGTGCCGAATATAG	1080
QY	1444	cttaacctcaactggaagttcccggaaggggaactlgaatagagaaaagctcccccactccac	1503
Db	1081	CTTACCCCTACCTGGAAGTTTCCGGGAAGGGACATGTATGAGAAAGCTCCCCATCTCCAC	1140
QY	1504	caaacaccttgctatagtactggtgtttaaagacaagcctccagaaatatcaatlatiagla	1563
Db	1141	CACACACCTTCTATAGTACTGTGTGTTATAGACAGAGCCTCAGAAATACAGATTTACTA	1200
QY	1564	cctgttataaacaagtggtgggaatgtaatactgggttaacccctgtgttttcaactca	1623
Db	1201	CTGTGTATTAACAGGTGGTGGGCTGCAATACATGCGGTATTAACCCCTGTGTTCACCTCA	1260
QY	1624	gtcttcaaccaatccaaagtgttggtgtatagtgccaatgtgtcccgagtgtaactac	1683
Db	1261	GTCTTCAACCAATCCAAAGATTTGTGTATGTGTCCAATGTGTCCCGAGTGTACTAC	1320
QY	1684	catcttgaggaagtggtcctctatgaaataagcatatcggtlataaacgaccaaagaagaga	1743
Db	1321	CATCTGTAGGAAGTGTCCTTGATGATATAGCATTCGGTATTAACCGACCAAAAGAGAA	1380
QY	1744	cccgatattccctacctaagctgtatagtcgcgatctagggagcggccgttggcgtlaggaaca	1803

QY 2101 gaagaagagctgaccaggggtggttgaagaagatggttcaacaggtctccttgatgacc 2160
 Db 7836 GAAAGAGAGGCTGACACAGGGGTGTTTAAAGATGTGTCAACAGGTCTCTTGATGACC 7895
 QY 2161 accggtctctgctcgaagggcccttagtccgtccgttacttaacagttggg 2220
 Db 7896 ACCGCTCTTCTGCTCTACGAGGACCCCTAGTACTCCGCTCCGTTACCTTAAGTTGG 7955
 QY 2221 ccttgcttaataataggttgttgccttcttgtagaagacagtgatgacagcagatc 2280
 Db 7956 CTTGCTTAATTAATAGTGTGTGCTTTGTGTAGAGACAGATGATGACATC 8015
 QY 2281 atgttacttaggcaacagttacaaagcctcttgagccaagagaaactgaacttagct 2340
 Db 8016 ATGCTACTTAGGCAACAGTACCAAGGCTTCTGAGCCAGAGAACTGACCTTACCT 8075
 QY 2341 tcccaagttctaagttagaactataacagacaagaagtggggaatgaaagatgaaaa 2400
 Db 8076 TCCCAAGTTCTAGATTGAACTATTAAACAAGACAGAAAGTGGGAATGAAGATGAAA 8135
 QY 2401 tgcacacttaacccctccagaaccagaagaatlaataaaagctctaagtccccgaatt 2460
 Db 8136 TGCAACCTAACCCCTCCAGAACCCAGAGAGTTTAAAGCTTAATGCCCCCAATT 8195
 QY 2461 acaagaccctgctgctgctcgaatgaatgagtaagagtcacacttccatgttccagggc 2520
 Db 8196 CCAAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8255
 QY 2521 ctgctatccctgctcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2580
 Db 8256 CTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8315
 QY 2581 ctgttaaacctgactgagcacaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2640
 Db 8316 CTGTAAACCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8375
 QY 2641 tcccaactgactgactgactgactgactgactgactgactgactgactgactgactgact 2700
 Db 8376 TCTCAACTGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8435
 QY 2701 ttttaaatgactgactgactgactgactgactgactgactgactgactgactgactgact 2760
 Db 8436 TTTTAAATGATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8495
 QY 2761 gaggcgaggtctcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2820
 Db 8496 GAGCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8555
 QY 2821 ccccaataaagctgctcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2880
 Db 8556 CCCCAATAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8615
 QY 2881 taagactgtgggcccagcgctgtgaataaaatcctctgtctgttgcataa 2936
 Db 8616 TACGACTGTGGGCCCCAGGCGCTTGGAATAAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 8671

RESULT 5
 LOCUS AX002804 3482 bp DNA
 DEFINITION Sequence 3 from Patent WO9853104.
 VERSION AX002804
 KEYWORDS AX002804.1 GI:9885132
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 Porcine endogenous retrovirus
 Viruses; Retroviral viruses; Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 Stoye, J.P. and Weiss, R.A.
 Detection of retroviral subtypes based upon envelope specific
 sequences
 JOURNAL Patent: WO 9853104-A 3 26-NOV-1998;

MEDICAL RES COUNCIL (GB); STOYE JONATHAN PAUL (GB)
 FEATURES
 Source
 1.3482
 /organism="porcine endogenous retrovirus"
 /db_xref="taxon:61673"
 BASE COUNT 927 a 854 c 867 g 834 t
 ORIGIN
 Query Match 59.0%; Score 1743; DB 6; Length 3482;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 2913; Conservative 0; Mismatches 21; Inbels 1; Gaps 1;
 QY 1 tggctttaggtttagaagaccccttgagcaatttgggtgacccctatgattgctct 60
 Db 548 TGGCTTTTAGGGTTAGGAACCCCTGAGCAATTTGGGCTGACCCCTATGAAATGCTCT 607
 QY 61 acgggggaacccccccgttggtagaattgtctctgataagttgtatgtgctctt 120
 Db 608 ACGGGGACCCCCCGTTGGTGAATGCTTCTGTACATAGTCTGATGTGCTGCTTT 667
 QY 121 cccagccttctctctaggtcgaagcgtcgtggttgggtgagcaagcgttgaagc 180
 Db 668 CCCAGCCTCTGTTCTTAGGCTCAAGGCGCTGAGTGGGTAGGCAAGAGCTGGAAGC 727
 QY 181 agctccggagagccttactcagaagagagagacttgcagttccacatcgctccaagttg 240
 Db 728 AACTCCGGGAGGCTCTACTGAGGAAGAGACTTGCAGATGCTTCCAAATGCTG 787
 QY 241 gagattcagctatgtttagacgcacacgttgaagaaacctggaactggttgaagagac 300
 Db 788 GAGATTCAAGTCTATGTAGAGCCGACCTGCAAGAACTCTGAGACTCGGTGGAAGGACC 847
 QY 301 ctatctcgttactttagac 360
 Db 848 CTATCTGCTGCTTGTGACCAAC 907
 QY 361 tccatgcaatccacagcttgaagcttggcaccctccgactcgggttgaagagcgaagaaga 420
 Db 908 TCCATGCAATCCACAGCTTAAGCTGGGCGACCTCCGACTGGGTTGAGAGGCCCAAGAA 967
 QY 421 ctgagaatcccttaagcttgcctcactgcctcgttgccttacttcaacaaatgaccc 480
 Db 968 CTGAGAAATCCCTTAAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
 QY 481 caggccagtagtaaacgcttatagacagctcgaaaccccatagaccttatccctacc 540
 Db 1028 CAGGCGAGTAAACGCTTATAGACAGCTCGAAACCCCATAGACCTTATCCCTTACC 1087
 QY 541 tggctgattatgacccctgatacgggtgtcactgtgataatgacactcgaagttgttcct 600
 Db 1088 TGGCTGATTATGACCCCTGATACGGGTGTCACTGTAAATGACACTGAGGTGTGCTCT 1147
 QY 601 agaggaacctggttggcctgaactgcatcttgcctcgaattgataaccgctgtlaaa 660
 Db 1148 AGAGGCACTGTGGGCTCTGAATGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
 QY 661 agcacacctcccaacacagcgttagttatggtttctatgtcgtgcccagagcagaaga 720
 Db 1208 AGCACACCTCCCAACCTGAGTCCGATGATGAGGTTCTATTTGCTGCGCCAGACAGAGAA 1267
 QY 721 gagaataactgttgggttcttgggaatccctgtttagagagatgaggtgtgctcacctcc 780
 Db 1268 GAGAAATACTGTGGGGTTCTGGGGAATCTTCTTGAAGATGAGAGTGTGCTACCTCC 1327
 QY 781 aagagatgagacttgaatgagccatctcctcagaagccgggtlaaatcttctctgtc 840
 Db 1328 AACGATGGAACCTGGAAGAGGCGGATCTCTCCAGACCGGGTAAATTTCTCTTGGC 1387
 QY 841 aattccggccgggcaagtaacaaatgataactataaagataagagcgtgtcccca 900
 Db 1388 AATTCCGGCCGGGCAAGTACAAAGTATGAAATATATTAAGATTAAGAGCTGCTCCCA 1447

QY	1501	caccacaaccttgctatagctacgctcgtggtttatagagaagcgctccagaataacgaattta	1560
Db	2048	CACCACACACCTTGGCTATAGTACTGTGGTTTATAGACAGCGCTCCAGAAATACGATTTTA	2107
QY	1561	gtacccttgattataaagagcgtgctgagcgtacgaataactggttlaacccctctgtctccacc	1620
Db	2108	GTACCTTGTTATTAACAGGTGGTGGGCAATGCAAACTACTGGGTAAACCCCTGTGTTTCCACC	2167
QY	1621	tcagttctcaaccaatccaaagaatttgtgtgcatggtccaaatcgctcccccgaagtac	1680
Db	2168	TCAGTCTTTCACCAACCAATCCAAAATTTCTGTGTGTCATAGTGCCAAATCGTCCCCAGATGTAC	2227
QY	1681	tacacaccttaggaagatgtgtcccttatataatactgacactcgtaataacgcagcaaaaaaga	1740
Db	2228	TACCATCTTAAGGAAGTGGTCTTGATGATGATGTGACTGCGTATAACCAACCAAAAAAGA	2287
QY	1741	gaacccgctatcccttaaccctagctgtaatgctcggataaggaagcgcgtgtgctgaagaa	1800
Db	2288	GAACCCGATACCCCTTACCCCTAGCTGTAAATGCTCGGATTATAGGACGCGCGTGTGGCGTAAAG	2347
QY	1801	acagaggaacgtcgcctcgatccaaaggaaccaagcagcgtcagaagaagaagactgtgtgaacta	1860
Db	2348	ACAGAGGACGCTGCCCTGATCCACAGGACCAACAGCAGCTTAGAGAAAGACTTTGGTGAAGCTA	2407
QY	1861	catgcgcgcattgacaagaagatctccagagccttlaaagagagctgctatgaacactgaagaag	1920
Db	2408	CATGCGGCGCATGACAGAAAGATCTCCAGGCTTAGAGAGATCGTTAGCAACCTTGAAGAG	2467
QY	1921	tcacctgactcttctgtctgaagtgtgtctctacagaaacccgagaggaattagatctctgttt	1980
Db	2468	TCCCTGACCTTCTTGTGCTGTGAAGTGTTCTTACGAACCCGAGGGGATTAGATCTGCTGTTT	2527
QY	1981	ctaaagaagaagctgggtctatgtctgaagccttlaaagaagaatgtgtctctatagtatgatac	2040
Db	2528	CTAAGAGAAAGTGGGTTATGTGCAAGCCTTAAAGAAAGATCTGCTCTATGATGATCAC	2587
QY	2041	tcagaagacatcaagaagactccatgaaacaagcttagaaaaagcttagaagagcgtccgaag	2100
Db	2588	TCAGAGGCGCATCAGAGACTCCATGACAGAGCTTAGAGAAAGGTTTAGAAGGCGTCGAGAG	2647
QY	2101	gaaaagaagagcttgacccagaggggtgtgttgaagagatgttcaaaagcttcccttgatgac	2160
Db	2648	GAAAGAGAGCGGTGACCGAGGGGTGGTTTGAAGAGATGTTTAACAGGTCTCCTTGATGATCC	2707
QY	2161	accgcgccttctgctctgaagcggggccctctagctgctcgtcttacttaacagttgag	2220
Db	2708	ACCCTGCTTTCGTGCTGTACGGGACCCCTAGTGTGCTCTGCTCTGTTACTTACAGTTGGG	2767
QY	2221	ccgtgcttaataataggttgtgtgcttgttaagaaacgggtgaagagagctccaagtc	2280
Db	2768	CCGTGCTTAATTAATTAAGTGTGTTGTCCTTTGTTAGGAACGATGTGCTGCATGCTCAGATC	2827
QY	2281	atggtgaacttagcaacagctaccagaagcctctgagccaaagagaaactgaacctctagct	2340
Db	2828	ATGGTGACTTAGCAACACTACCAAGGCGCTTCTAGGCCAAGAGAAACTGACCTTAGCGCT	2887
QY	2341	tcaccagttcctaagattagaactatlaaacaagaagaagtggtggaaatgaagaagtgaata	2400
Db	2888	TCCCAATTTTATAGATTGAACATTATTAAACAAGACAGAAAGTGGGCAATGAAGAGATGAAGA	2947
QY	2401	tgaacacttaacctcccgaaacccaggaagatataaaagcctctaaatggtccccgaatt	2460
Db	2948	TGGAACCTTAACCTCCCGAAACCAAGAGAGTTATTAAGAAAGCTCTAAATGCCCGGANTT	3007
QY	2461	acagaccctcgtcgtctgcaglaaataagtagaagagtcacacttccattgttccaagtc	2520
Db	3008	CCAGACCTTGCTGGCGCCAGTAATATAGTGTGAAGGTCACACTTCTATTGTTCCAGGCG	3067
QY	2521	cgtctactcgtgacctgaagatgaataaaggaagaatgaattgaactacggtctctcgtgatt	2580
Db	3068	CTGCTTCTCTGGCGCTTAAGTAATTAACGGAATGAAGATGTACTGACTTAATCGCTTATCGANTT	3127

Qy	2561	ctgtaaactgaactgagccacatagaagaatttgattacaatttgcacggccttagtaccta	2640
Dd	3128	CTGTAAATGATTGGTGCACAGCAGCGCGGCTTCGATATTTTAAAAVGAATTGGTCCACG	3187
Qy	2641	tctaacatgcgaactgtgtaactctgtcccaggagaccacagacatgcygagacctccgagcta	2700
Dd	3188	TCTCAACTGCATATCTGTACTCTGCCAGGAGGCCACGAGATGGGAGACCTCCGAGACTA	3247
Qy	2701	ttttaaatgatctgtgccagcgagcgcgagctcgcgatltttaaatgatctgtgtccatg	2760
Dd	3248	TTTTAAATGATTGGTGCACAGCAGCGCGGCTTCGATATTTTAAAAVGAATTGGTCCACG	3307
Qy	2761	gagcgcgagcttc-tcgatatltttaaatgatctgtgtgtaacgacagagcttgcgtga	2819
Dd	3308	GAGCGCGGCTTCGATATTTTAAAAVGAATTGGTGTGTGACGCACAGCCTTGTTGTGA	3367
Qy	2820	accaccaataaagctgtcccgatctccgcactcgtggcgccgagctcctaccctgagtgt	2879
Dd	3368	ACCCCATATAAAGCTGTCCCGATTCGCGACTCGGGGCGCGAGATCCTCACCCCTCGGTGT	3427
Qy	2880	gtacgactgtggcgccccagcgagcgttgtgaataaaaatcctctgtcgttttgatc	2934
Dd	3428	GTACGACTGTGGCCCCAGCGCGCTTGGAATTAATAATCTCTGCTGTGTTCATC	3482
RESULT	7		
LOCUS	PERY17013	7808 bp	DNA
DEFINITION	porcine endogenous retrovirus gag, pol, env genes.	VRL	14-APR-2000
ACCESSION	Y17013		
VERSION	Y17013.1	GI:3005985	
KEYWORDS	env gene; gag gene; pol gene. porcine endogenous retrovirus. porcine endogenous retrovirus viruses; Retroid Viruses; Retroviridae; Mammalian type C retroviruses; I-Mammalian type C virus group.		
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Czaderna,F., Fischer,N., Boller,K., Kurth,R. and Tonjes,R.R.		
TITLE	Establishment and characterization of molecular clones of porcine endogenous retroviruses replicating on human cells		
JOURNAL	J. Virol. 74 (9), 4028-4038 (2000)		
MEDLINE	20219394		
REFERENCE	2 (bases 1 to 7808)		
AUTHORS	Tonjes,R.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-MAR-1998) R.R. Tonjes, Paul-Ehrlich Institut, Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG		
FEATURES			
source	Location/Qualifiers		
	1..7808		
	/organism="porcine endogenous retrovirus"		
	/db_xref="taxon:61673"		
	/tissue_type="kidney"		
	/cell_line="PK15"		
	260..1834		
gene	/gene="gag"		
	260..1834		
CDS	/gene="gag"		
	/codon_start=1		
	/protein_id="CAA76581.1"		
	/db_xref="GI:3005986"		
	/translation="NGQVTTPPLSLTDHMTTEVRSAHNLSVOYKKGPMOTFCASEMPN ITDVGWPSGETFNSSETILAVKAIRIOTEPSSHPIDDEPIITLMOLADLPMPVPYMLN KRKKGPRIILTAGENKKSIAEKVEVSPTIPEIEPPWPBPQCVAPPPIPYAOGAVSG CSAPGPAPVPEGPAAGTRSRGARTEPERIDEALIIPLRTYGVPMFGQLPLOTWPFSS ADLYNNKTNNPFESDPORLTGLVESLMSFSQPDWDQQLQLTLFAEERILLDA RLRMGDADRPTLOJNEIDMCPFLTRIMDYNTAGRESLKITYROALVAIGASRRRP TNLAUKREVMQGNPNPEPSVFLELRMEAFRRFPETPTSEAOAKASYALAFITGSALDIR KI.KIQRIEGIOEAELRLVLYAEAKAYGGRTEBEKOREKEREEEREERDRROKNLLT KIIAAVEGRSSRRBRDFRKIRSGPROSGNLGNTPILDQCACKEKGMANNCKK KGNKGPKVIALLEDKND"		
	2120..4726		
	/gene="pol"		

CDS

2120..4726
/gene="pol"
/codon_start=1
/protein_id="CAA76582.1"
/db_xref="GI:3005967"
/db_xref="SPTREMBL:O73505"
/translation="MGAQISFEDGKREVSNNKPTVTILTLQIDDEYRUYSPQVKFPDQ
IQSMLEDFPQAMAEATAGMGLAKOVPOVITOLASATPVSVRQPLSRAREIMPHVQ
RLIQGILVNVQSPMPTPLPVKPGSTNPRKVDLEENKRVODIHPVLEIMPHVQ
ALPPEKMTVVDLKDAPFCLRLHPTSPDLFAFEMRDEPQSGTGLMTRIPQGNKS
PTIFEDALHDLNFRIOHPVITLQYVDLLAATQDCEGKALLLELSDGTR
ASAKKAOICREVTYGLSRGGRVLTARAKTYVOIPAPTTAKOVEFGLTGFYR
LMTFGPATLAAPLYPTKEKESPMAPBOKAFDAIKALAPALAPDVKFPTLY
VDERKVAVALVLTQIGPWRPVAVLSKLLDVASGMPVCLKALIAVALIVKADKL
LGNITVIAHALENIVRQPDMMTNAMTHYOSILLETERTAPALNPATLLPE
ETDEPVDHCHOLLLEETGRKDLIDPLTGEVLTMTDGSYVEGRMGAADVOC
THTIMASSLEPESAOAKELMALTOALRAEKSINITYDSRYAPATPHVGAITYKOR
GLTISAGREIKNKEETLSLEALHLPKRLATIHCHGKAKADLSRGNOMARVAKO
AONVNLPTIEMPRAPERROTYLLEDMOETIKIDQSFETPSTGYSKIDLPKES
LEYVOQIHLLTHGTRKLDQLVKISPYHVLRLPGVADVVAHVCPCOLVANNPSRIP
GKRLRSHPGAHMEVDTEVKPAKYNKYLIVFDVTFSGWEAVPTKETSTVAKKY
WRKFFRDLEYLR
5294..7267
/gene="env"
5294..7267
/gene="env"
/codon_start=1
/protein_id="CAA76583.1"
/db_xref="GI:3005988"
/db_xref="SPTREMBL:O73506"
/translation="MHPTLSMRHLPTRGGEPRRLRIPISPAIWFLLTITTPQASK
RLIDSNPRLSLTWLITDIDPGVYVSTGVAIPRGTMWPELHFCRLINPAKSTP
PILVSYGCGCTEKEKRYCGSESCRCWSCVTSNDGKWPISLQDRKFSNP
SGPQKRYKWLTKDKSCSPSDLYIKISTEKGQENOMKINMSGIVIFYKXGA
GSTLIRLKTETGEPVAVGPDVLAQGPALPEPMLPVLDTLRDIPQPSN
GTTGLIPNTPRNSPEVPVATGKORFLSLIOGAPALINSTPDATSCMLSSPPY
EGMAKEGFNVYKHEKNOCTWGSNNKLLTLEVSQKICIGAPASHOHLKSYVVEO
ASENOYLIVGYRMWACNTGLPCVSSSVENOSDPCVMQIVRYVHYHBEVYLD
DYRNRPREPVSILTAVMLIGTAVYCGCTAALTITGPOLKEGLDELHAAETEDR
ALEESVNSLESTLSISVYIQNRGDLTLRLREGGLCALKECCCVYHSAIRNS
MSKLERLERERREHEDQMFEGMFPNSPMWTLTSLATGLPVLVLLLTLPVCILIN
RFVAFVRKVAVOIMVLRLQYQGLISQGETDL"
BASE COUNT 2077 a 1971 c 2047 g 1713 t
ORIGIN

Query Match 56.0%; Score 1654; DB 14; Length 7808;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2854; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 tgcatttagagtggaacaccccttgacagtttggtgctacccctatgaattgctct 60
DB 4931 TGCATTTAGGTTAGGAACACCCCTTGACAGTTTGGGCTACCCCTATGAATTGCTCT 4990

QY 61 acgggggagaccccccgttgtagaattgctctgctacatagctgtagtgctgctt 120
DB 4991 ACGGGGGAGACCCCCCGTTGGTAGAATTGCTCTGTCATAGTGAAGTGTGCTGCTT 5050

QY 121 cccagcccttftctctgagctcaagcgctcgagtggtgtagggcaacgagcgttgagc 180
DB 5051 CCCAGCCCTTFTCTCTGAGCTCAAGCGCTCGAGTGGTGAGCAACGAGCGGTGAGAC 5110

QY 181 agctccggagagcctactcaagagagagagagcttgcaagtctccacatcgcttcaagt 240
DB 5111 AGCTCCGGAGAGCCTACTCAGAGAGAGAGAGCTTGCAAGTCTCCACATCGCTTCCAGAGTG 5170

QY 241 gagatcaagctatgtagagcgacacgctgcaagaaacccctgagactcggttggaaggagc 300
DB 5171 GAGATTCAGCTATGTTAGACCGCACCGTGAGAAACCTCGAGACTCGGTGGAAGGCGC 5230

QY 301 ctatctcgtaactttgacacacacacgctgtggaagtcggaagaaatctccacttga 360
DB 5231 CTATCTCGTACTTTTGACACACACACGCGTGTGAAGTGAAGAACTCTCCACCTCGGA 5290

QY 361 tccatgcatcccaagcttaagctgagccacccctccgaactcggggtagagagccgaaaga 420
DB 5291 TCCATGCAATCCCAAGCTTAAGCTGAGCCGACCTCCGACTCGGGGTGAGAGCGGAAGA 5350

QY 421 ctggaatccccccttaagcttcgacctcgaatcgccctggtcttcccttaacaataactccc 480
DB 5351 CTGGAATCCCCCTTAAGCTTCGACCTCGACCTCGCTGCTGCTTAATCAATAACTCCG 5410

QY 481 caggccagtagtaagagcccttataagagctcgaaaccccatagaccttaccctacc 540
DB 5411 CAGGCCAGTAGTAAGAGCCCTTATAGACAGCTCGAACCCCATAGACCTTATCCCTTACC 5470

QY 541 tggcttgattttagaaccttgtagcggtgctacgtgtaataagcaactcgagtgctcctc 600
DB 5471 TGGCTGATTTATGACCTTGATGAGGAGTGCATCGTAAATAGCACTCGAGCTTCTCCT 5530

QY 601 agaggccactggtgagcctgagacgcaatctctgacctcgatlgattaaacccgctgttaa 660
DB 5531 AGAGGCACTGAGGAGGCTGAGACGATTTCTGCTCGATTTATTAACCCGCTGTAA 5590

QY 661 agcaacactcccaacactagtcgtagttagtggtctatgctgcccagagagagaa 720
DB 5591 AGCAACACTCCCAACACTAGTCGCTAGTATGGGTTCTATTGCTGCCAGGACAGAGAA 5650

QY 721 gagaataacttgagggtgctcggagagatctctctctgtagagagtagagctgtaacctcc 780
DB 5651 GAGAAATACTGTGGGGGTTCTGAGGATCTTCTGTAGAGATGAGAGTGGCTGCTACCTCC 5710

QY 781 aacgataagagactggaatagcgacatctctccagagacgggtaaaatctcctgtg 840
DB 5711 AACGATAGAGACTGGAATAGGCGCATCTCTCCAGAGACCGGTAATTTCTCCTTGTG 5770

QY 841 aattccggccggggagagaaatagatgaactatataagaataagagctgctcccca 900
DB 5771 AATTCCGGCCGGGAGAGAACAAAGATGAAGAAATATTAAGATMAACGCTGCTCCCA 5830

QY 901 tcagacttagataltatcctaaagaataagttcactgtaaaagagaaacaggaatattcaa 960
DB 5831 TCAGACTTAGATATATCTAAAGATTAAGTTCACTGAAAAAGGAAAAAGAAATATTC 5890

QY 961 aagtgataataatgtagatagcgtggggaatagtttttaataatatgaggggagagag 1020
DB 5891 AAGTGATTAATATGATATGAGCTGGGGAATAGTTTATTAATATGCGGGGAGCAGG 5950

QY 1021 tccacttaacatccgaccttagagtagagagcgggagaaacccctgtagagatgga 1080
DB 5951 TCCACTTAAACATTCGCTTAGAGTAAAGAGAGGAGGAGAACCCCTGTGCGAGTGGGA 6010

QY 1081 cccgataaagtaactggtctgaacaggggccccgggacctgtaggacacgataacttgcg 1140
DB 6011 CCCGATAAAGTACTGCTGAGAACAGGGGCCCCGGGCGCTGAGGACCGGATTAATCTG 6070

QY 1141 gtagcccaatlaactcgtgctgctgagctgagcaatacaacagcgctagcaacaglacact 1200
DB 6071 GTAGCCCAATTAACCTGCTGCGGCTGATCAATAACAGCGGCTAGCAACGCTTACACT 6130

QY 1201 ggaattgattcctacaaacagccttagaaaccccaagggttctcgtttaagaagaagag 1260
DB 6131 GGATTGATTCTACCAACAGCGCTAGAAACCTCCCAAGTGTCTCTGTTAAGACAGAGACG 6190

QY 1261 agaccttcagctctcatccagggagcttccaagcctcaactcaactccagacccctgtg 1320
DB 6191 AGACCTTCACTCTCATCCAGGAGGCTTCCCAACCATCAACATCCACCGACCTGATGCG 6250

QY 1321 acttctcttggttggttcttctctcctcctcctcctcctcctcctcctcctcctcctc 1380
DB 6251 ACTTCTTTTGTGCTTGTGCTTATCTCTCAAGGCGCTCTCTTATTAAGAGGAGTGTCTTAA 6310

QY 1381 gaaagaaatcaaatgtgacaaagagacagaaatcaatgaatgtagaggggtccggaat 1440
DB 6311 GAAGGAAATTCATATGTGACCAAGAGACATGAAATCAATGTACATGAGGGGTCCGAAAT 6370

QY 1441 aagcttaccctcactgaagttccgggaaggagacatagatagaagaagctcccccaccc 1500
 DB 6371 AAGCTTACCCTCCTGAGTTTCCGGGAGGAGACATGATAGGAAGACCTCCCATATCC 6430
 QY 1501 caaccaaccccttgctatagtagtctgtgtttatagagagcctcagaataatcatattta 1560
 DB 6431 CACCAACACCTTGTCTATAGTACTGTGTATGACACAGCCCTCAGAAATAGATATTTA 6490
 QY 1561 gtaccggttatataagagtggttggagatcaactagggtttaacccccgtgttccacc 1620
 DB 6491 GTACCTGGTATATACAGGTGGTGGCATCAATCTGGGTTAACCCCTGTGTCTTCTCC 6550
 QY 1621 tcaagttcaaccaatccaaagaattgltgtlcatagtlccaaatcgtcccccagatgtac 1680
 DB 6551 TCAGTCTTCAACCAATCCAAAGATTCTGTGTATGTGTCTCAATATGCTCCCGAGATGAC 6610
 QY 1681 taccatcttgaggaagtggtccttgatgaatactactatcgatataaccgacccaaaga 1740
 DB 6611 TACCATCTGAGGAAGTGGCTCTGTATGATATGACTATCGGTATMACCAGCAAAAGA 6670
 QY 1741 gaaccggtatccctaccctagctgtaatgctcgtatgagtgagagcggttggtgaga 1800
 DB 6671 GAACCCGATCCCTTACCCAGCTGTAATGCTCGATTAGGGAGCGCGTGGCGTAGGA 6730
 QY 1801 acagagacagctgccttgatcaacagacacagcaactagagaagaagacttgtagacta 1860
 DB 6731 ACAGGACAGCTGCCCTGATACAGAGACCAACAGCAGCTAGAGAAAGACTTGGTAGCTA 6790
 QY 1861 catgagccatgacagaagaatcgcagccttaagaagactgtgttagaacaactagaaga 1920
 DB 6791 CATGCGCCATGACAGAGATCTCCAGAGCTTGTAGAGAGTCTGTATGCAACCTAGAAAG 6850
 QY 1921 tccctactcttctgtctgaagtggttctacagaaccgagagagatagaatgtcgttt 1980
 DB 6851 TCCCTGACTTCTTGTGTGAAGTGTCTTACAGAACCGAGGGGATTAATCTGCTGTTT 6910
 QY 1981 ctaagaagaagtggtgtatgtcagccttaaaagaagaatgtgtcttcatgtatagacac 2040
 DB 6911 CTAAAGAGAGTGGTGTATGTGACAGCTTAAGAAAGAAATGTTGCTTATGTAGATCAC 6970
 QY 2041 tcaagagccatcagaagactcactgaacaagcttagaanaaaagtltagagaagcgctgaag 2100
 DB 6971 TCAGGAGCCATTCAGAGACTCCATGACCAAGCTTAGAGAAAGCTTAGAGCGCTCGAAG 7030
 QY 2101 gaagaagagctgacagagtggtgttagaagaagtgatcaaaagctcctgtatgagacc 2160
 DB 7031 GAAAGAGAGGCTGACCAAGGGGTGTTGAAGATGATGTTAAACGGGTCTCTTGGATGACC 7090
 QY 2161 accctgcttctgctcgtgaagggcccttagtagtctcgtcctgttacttcaagttgg 2220
 DB 7091 ACCCTGCTTCTGCTCTGAGGGGACCCCTAGTACTCTCTCTCTTACTTACAGTTGGG 7150
 QY 2221 ccttgcttaataataggttgttgccttlttagagaacagtagagtgagtcagacatc 2280
 DB 7151 CCTTGTCTTAAATATAGGTTGTGCTTGTAGAGAACGATGCGTGCAGTCCAGATC 7210
 QY 2281 atgtgcttagggaagaagtaaccaagccttctgacccaagagaagaactgaacctagacc 2340
 DB 7211 ATGTGACTTAGGCAACGATGCCAAGGCTTCTGAGCCAAGGAGAAACTGACCTTAGCC 7270
 QY 2341 tcccaagtcttaagattagaactttaacaagaagaagtgaggaaatgaagaagtataaa 2400
 DB 7271 TCCCAAGTTCTTAAGATTAGAACTATTAAACAAGACAGAGTGGGAGTGAAGAGATGAAA 7330
 QY 2401 tgaacactaacctcccagaacccagaagaattataaagaagctctaaatgtcccgcaatt 2460
 DB 7331 TGAACACTTAACCTCCAGAACCCAGGAAGTATTAATAAAGCTCTAATATGCCCCGAATT 7390
 QY 2461 acaagacctgtgtgagcagaatagtagagaagtgacaacttctctatgttccagggc 2520
 DB 7391 CCAAGACCTGTGCTGCTGCCAGTAAATAGTAGAAGGTACACACTTCTATTTGTTCCAGGG 7450
 QY 2521 ctgtctactcgtgacctagaatagaatgaagaagaatgagtgagtaactatcgttatt 2580

DB 7451 CTGCTATCTGCGCCCTAAAGTAAGATAACAGAAATGAGCTGACTATCGCTTATCTGATT 7510
 QY 2581 ctgtaaactgacttggacacacabaagaatgatattacacattgacagccagtagtacctta 2640
 DB 7511 CTGTAAACTGACTGGCACCATTAGAGATTGATTACATTTGACATTTGACAGCCCTTAGTACTTA 7570
 QY 2641 tctcaactgcaactcgtactctgtccagagacccacagatgcgagactcccgagacta 2700
 DB 7571 TCTCAACTGCAATCTGTACTCTGTCCAGAGAGCCACGACCAATCGAGACTCTCGAGACTA 7630
 QY 2701 tttaaaatgattggttccagagagcgagcgtctcgatattttaaatgattgtgcatg 2760
 DB 7631 TTTTAAATGATTGTTGTCACAGAGCGCGGCTCTCGATATTTTAAATGATTGTCTCACG 7690
 QY 2761 gagcgcgagcctcgtgatatlttaaatgattggttttgagcagcagagcttgttggaa 2820
 DB 7691 GAGCGCGGCTCTCGATATTTTAAATGATTGTTGTGACGACACAGGCTTGTGTGAA 7750
 QY 2821 ccccaataaagctgtccgcatlccgacactcgagcgagcagctctacccctgctg 2878
 DB 7751 CCCCATTAAGCTGTCCGATTCCGCACTCGGCGCGCAGTCTCTACCCCTGCTGTGG 7808
 RESULT 8
 PEN133816
 LOCUS PEN133816 8918 bp DNA VRL 07-DEC-2000
 DEFINITION Porcine endogenous retrovirus type C proviral gag, pol and env genes and LTR (class B, clone 33).
 ACCESSION AJ133816
 VERSION AJ133816.1 GI:6688945
 KEYWORDS env gene; env protein; gag gene; gag protein; pol gene; pol protein.
 SOURCE porcine endogenous retrovirus.
 ORGANISM porcine endogenous retrovirus.
 Viruses; Retroviridae; Mammalian type C retroviruses; 1-Mammalian type C virus group.
 REFERENCE 1 (bases 1 to 8918)
 Czauderna, F., Fischer, N., Bollner, K., Kurth, R. and Tonjes, R.R. Establishment and characterization of molecular clones of porcine endogenous retroviruses replicating on human cells
 JOURNAL J. Virol. 74 (9), 4028-4038 (2000)
 MEDLINE 20219394
 REFERENCE 2 (bases 1 to 8918)
 Tonjes, R.R. Direct Submission
 TITLE Submitted (04-MAY-1999) Tonjes R.R., Medical Biotechnology, Paul-Enrich-Institut, Paul-Enrich-Strasse 51-59, Hessen, D-63225 Langen, GERMANY
 JOURNAL
 FEATURES
 source Location/Qualifiers
 1..8918
 /organism="porcine endogenous retrovirus"
 /proviral
 /strain="type C"
 /db_xref="taxon:61673"
 /clone="33"
 /note="grown in PK15 (ATCC No. CCL-33) cell line"
 1..707
 545
 /note="cap site"
 710..727
 /note="91y4"
 1154..2725
 /gene="gag"
 1154..2725
 /gene="gag"
 /feature="gag"
 /function="retroviral capsid"
 /product="gag protein"
 join(1154..2725,2729..6313,6189..8162)
 /note="viral polypeptide"
 /codon_start=1
 /protein_id="CAB5339.1"
 /db_xref="GI:6688946"

/db_xref="SPTREMBL:0901X5"
/translation="WGQVTVPLSLTLDHTEYVRSANHLISVQKKKGMPTFCASWMP
TFDVGWSEGTENSELLIAKAIILFQTPGSHSDQEPYLLITWDLAEDPPWPKMLN
KPRKPEIRLILAEKNHSAEKVESPRIYELIEEPEPEPVPYPAQAGAS
PSAPGAPVEGAGAGRSRGAATPRTIDILALPTGPGPGGOLPDIOWPSS
ADLYNMKTNHPESDEPDLTGLVESIMESHOTMTDQCOLLOTLTTERRELLIEA
KRVNPGADGDPLOLNDIMGPRITGMDYNTAEORESLKIROLAVGLRASRSP
TNLAKEVWOGNEPNEPSVLEMLAEFRFTFPDTSEROKSAVALATGOSALDIR
KILRLGLELAEALRLVRAEKVYIRRETEERKREKEREERERERERERERER
KILAAVEKSSREERERIRSGRSGNLRNPLDEKQACVCKEYHARNAKPK
KGNKGVLTLEEDKDRGSDPLPEPRVYLKVEQVPELVYAGHSDLOPLK
KEKSNWMACTGROYWMTRTVDLAVGVTHSPFVTEPVGLGRLITRMKNOI
SFGQREPVSNKPIYVLITLDDERYLSPOVKDQDQSLBEFPOANATAMG
LAKQVPOVYLKASATPVSVROTISREKREPIWVORLQGLVPLVSDWNTPL
LPKREPTNDYRPVODLREYNKRVODIHPVBPENYMLALPERWYVLDLKDQF
CLRHLPTSOPLFAFEMWDEPQGTGOLTWRLPQGRKNSPTIDEALHMDLNFROH
POVTLQYVDDLLAGATKQDCEGKALLLELSDGRASAKKQICREVEYLYS
LRGQORLLEARKTYVOIPAPTTAKOVREPLGTAFRCIMIPGATLAPLYPLKE
KGFESNAPEHOKAPDAIKKALLSAPALAPDYTKPPTLYDEKGVARVLVOTLQPM
RRPVYLSKLDPYASGWPYCLAAVAILVADAKLIGONITVIAHALENITVRO
PPDMWMTNATHYQSLILTERVTEAPALNATILPEETDEPVHDLHOLETQ
VRDLNDILPTGVLTFTDSSVVEGRMAAAVVDGHTITWASLEGSAQAE
LMAITOLRLAEKGSINITYDSRYAFAHVAIKOKGLTASAREIKNEELISL
LEALHLPRKLAITHCPGHOKAKDLISGNOMADRVAKOAAVNLPIITETKAPBR
ROYTLEDMOERIKIDQESPEBCTYSKGLTLPHEGLLEVYQDTHRTHLGTRKLO
QIVRTSYHILRIPYADSVYKCHVPCOLYNANPSIPEGRKLGRSHGPAHMEVETE
VFAKTGNKILVAVDFTSGWVAEYPTKETSIVNAKILLETFRPFGPKYIGSNG
PAFAVOVSOGLALIGIDMKLHCAYRPOSSGOVERNARIKETILKLTETLINDIA
LPLPVLFRVNTPGQGLPYELLGPPPLVETASVADVLISQLEPRLKALEM
ROKAMKOLREAYSGEDLOVHFQYDGYVYRHAGNLETNAGKNSPVYLITTPAV
KVGASITWHTSHVKLAPPDGSMRAKEMENKILHLVLPYNNKNSPQOLHTPISW
RHLPTGRGEKRLRPLSPASIAMPLITITTPQASKRLIDSNRPRLSLTWLITIDP
DTGVYVNSTRGVAPRGTMPELHFCULINPAKSTPPMLVRSYGGCCPGKEKEXC
GDSGESFCRMSCVTSDNDGKWPISLQDRVKSFNVSNGPKWKLKDKSCPSD
LDLKLISFTEKQENIOKWINMSWGIYKXGGAGSTLILRLITETGTPPVAVG
PDVLAEOGPPLEPHNLPEPLTSLRDIOSPGGAGSTLILRLITETGTPPVAVG
GORFLSLIOGAFOAINTSDPDATSSCMLCSPPYEMAKMTCYKRNKNOCTW
GSRNKLITLIEVSKGKICAKPSHOLCTSYVYEOASBNOLYVGYRMAKNTGL
TPCVSTSVENOSKDFCVMOYIPRYVTHEVEVLDYDRIYNPKREPVSLTAAVGL
LGTAVGVTGTAALITGPQOLEGELHAAEMEDIRALEESNIEESLISLSEVL
QNRNGIDLFLREGGLCAALKKECCFYVHSGAIRMSKLEERLERERREEDOGW
FEGMFRSPMTLLSALTGPLVLLVLLLVGCLINRVAVFREVSAVOJMLVRQO
YOSILSGSERDML
2729. 6313
/gene="pol"
/function="retroviral protease and polymerase"
/product="pol protein"
2729. 6313
/gene="pol"
6189. 8162
mat_peptide
/gene="env"
/function="retroviral envelope"
/product="Env protein"
6189. 8162
gene
/gene="env"
8212. 8918
LTR
8811. 8816
polyA_signal
2337 a 2245 c 2334 g 2012 t
BASE COUNT
ORIGIN

Query Match 55.3%; Score 1636; DB 14; Length 8918;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2736; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 tgccttttaggttagaacaccccttgagacgtttggagccctcctatgaattcctc 60
Db 5826 Tgccttttaggttagaacaccccttgagacgtttggagccctcctatgaattcctc 5885
QY 61 acgagggagaccccccgttgtagaaatctctctgacacatgagtgatgagtcctt 120
Db 5886 ACgagggagaccccccgttgtagaaatctctctgacacatgagtgatgagtcctt 5945

QY 121 cccagccttctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180
Db 5946 CCCAGCCT 6005
QY 181 agctccggagagcctcctcctcctcctcctcctcctcctcctcctcctcctc 240
Db 6006 AGCTCCGGAGGCGCTCTCTAGAGGAGAGAGACTTGAAGTTCCACATCTCTCAAGTGG 6065
QY 241 gagatcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 300
Db 6066 GAGATTCAGTATGTTAGACCGCACCGTGCAGAGAAACCTCGAGACTCGTGGAAGGGCC 6125
QY 301 cttatctcgtactcttgcaccacccacccacccacccacccacccacccaccc 360
Db 6126 CTTATCTCTGACTTTTGACACACACCAACCGCTGTGAAGTCCAGAAATCTCCACTGGA 6185
QY 361 tccatcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 420
Db 6186 TCCATACATCCACGTTAAGCTGCGCCACCTCCGACTCGGGGTGGAGCGGAAAGGA 6245
QY 421 ctgagatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 480
Db 6246 CTGAGATCCCTTAACTTTCGCTCCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6305
QY 481 cagggcagctagtagaacgctctatagacagctcgaacccacccacccacccaccc 540
Db 6306 CAGGCGAGTAGTAAGACGCTTATAGACACTGGAACCCCATAGACCTTATCCCTTAC 6365
QY 541 tgcctatattagacccctcctcctcctcctcctcctcctcctcctcctcctcctcctc 600
Db 6366 TGCTCTATTATGACCTCTATACGGGTGTCACCTGTAATAGCACTCGAGTGTCTCTCT 6425
QY 601 agagacacccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 660
Db 6426 AGAGGACACTCGGAGGCTGACACTGCAATTCGCTCCGATGATTAACCCGCTTTAA 6485
QY 661 agcagacccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 720
Db 6486 AGCAGACCTCCCAACCTAGCTCCGTATAGGTTCTATGCTGCCAGGCGCAAGAGAA 6545
QY 721 gagaataactctgaggggctcctcctcctcctcctcctcctcctcctcctcctcctc 780
Db 6546 GAGAAATFACTGTGGGAGATCTG6GGATCTCTCTAGGAGATGAGACTGCTCACCTCC 6605
QY 781 aacgagtagagacttggaatcctcctcctcctcctcctcctcctcctcctcctcctc 840
Db 6606 AACGATGAGACTGGAATGAGGCGATCTCTCCAGGACCGGTAATTTCTCTTGTGTC 6665
QY 841 aattccgagccgagcagtagaataatgtagaataatgtagaataatgtagaataatgtaga 900
Db 6666 AATTCCGCGCCGCGCAAGTACAAAGTATGAAATATTAAGATTAAGAGCTGCTCCCA 6725
QY 901 tcaagcttagattactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 960
Db 6726 TCAGACTTAGATTATCTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 6785
QY 961 aagtgataaactgtagcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1020
Db 6786 AAGTGATTAATGATATGATAGCTGGGGAATAGTTTATTAATATAGGCGGAGACAGG 6845
QY 1021 tccacttaacactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1080
Db 6846 TCCACTTAAACATTCGCTTATGATATGATAGAGAGGAGAGAGAGAGAGAGAGAGAGAG 6905
QY 1081 cccgataaagtagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1140
Db 6906 CCCGATTAAGTACTGCTGTAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6965
QY 1141 gtcgccaataactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1200
Db 6966 GTGCCCAATTAACCTCGTGGGCTGACATTAACACAGTGCCTTACGACAGGATGACACT 7025
QY 1201 ggaatgattcctacacacagcctagaaactccacagtgctcctcctcctcctcctcctc 1260

Dp	7026	GGATTATTTCTTAACCAACAGCGCTTGAAGAACTCCCAAGTGTTCTGTTAAGACAGGACAG	7085
Qy	1261	agaacttcagtcctccatccaggagacttccaagccatcaactccaccgacctgatygc	1320
Dp	7086	AGACTCTTCAGTCTCATCCAGGAGACTTTCACAAGCATCAACTCCACCGACCTGATGACC	7145
Qy	1321	actctctctgttggctttgtctatccccaaggacctcttlatagagggatgagctcaaa	1380
Dp	7146	ACTTCTCTTGTGTGGCTTGTGTCTTAACCTCAGGCGCTCTTATTATGAGGGGATGGCTAAA	7205
Qy	1381	gaagaagaatccaatgtgacccaagaagatagaatacaatgatactgaggggtccggaat	1440
Dp	7206	GAGGAAAAATTCATATGTGACCAAGAGCGATGAAATCAATGTACATGGGGGTCCCGAAAT	7265
Qy	1441	aagcttaacctcaatgaagattccgggaaagggaatgataagaaaagtccccaatc	1500
Dp	7266	AAAGCTTAACCTCACTCAAAATTTCCGGAAAGGGGACATCATAGGAAAAAGCTCCCAATCC	7325
Qy	1501	caccaaacacctgtgcatagtactgtgttlatagacagacctagaanaatcaatlatla	1560
Dp	7326	CACCAACACCTTTGCTATATGTACTGTGGTTTATGACGAGGCTCAGAAAATCATATTTA	7385
Qy	1561	gtacctgttataacaagtgatggtggacatgacaatctgggttaacccccgtgttccac	1620
Dp	7386	GTACCTGTGTTATTAACAGGAGTGGGGACATCAATTACTGGGGTTAAACCCCTGTGTTTCCACC	7445
Qy	1621	tcagcttcaaccaatccaagaatttgatgtatcagtcgacaaatgcgtcccccgatgac	1680
Dp	7446	TCAGTCTTCAACCAATCCAAAGATTTCGTGTGATGTCGCAATTCGTCCCGAGGTGAC	7505
Qy	1681	taccatccctaaagaaatggttcccttgatgaataatgactacgtataaccgacccaanaa	1740
Dp	7506	TACATCTCTGAAGAAAGTGAGCTTGTATGTAATGTACTATCGGTATTAACCGACCAAAAAGA	7565
Qy	1741	gaaccggtatcccttaacctagctgtatgaatgctcggaattagaggacggccgtgtggatga	1800
Dp	7566	GAACCCCTTAATCCCTTAACCTTAAGCTGTAAAGTCTGGATTTAAGGACCGCGCTGTGGAGTGA	7625
Qy	1801	acaggagacagctgcacctgatacacaagaaccaacagcagcttagagaagaagacttgtagctla	1860
Dp	7626	ACAGAGCAACAGTCCCTGATCACAAGGACCAAGACAGACTAAGAAAGGACTTGGTAGCTA	7685
Qy	1861	catgctggcatatgacagaagaatctccagaccttaaaagagttctgttagcaacctagaagag	1920
Dp	7686	CATCGCGCATATGACAGAAAGATCTCCAGCGCTTAAGAGAGATCTGTTAAGAACCTTGAAGAAAG	7745
Qy	1921	tcacctgacctcttgatctgtagaagtgatctcaagaacccgaaggagattagatctcgtgtt	1980
Dp	7746	TCCCTGACTCTTTGTGTGAAGTGCTTTACACAACCGSAGGGGATTAATCTCGTGT	7805
Qy	1981	ctaagaagaagtggtgtatgtgtcagaccttaaaagaagaatgttgtctctatgtatgaatc	2040
Dp	7806	CTAAGAGAAGGTGGGTATGTATGACAGCGCTTAAGAAAGAAATGTTGCTTATATGTAGATCAAC	7865
Qy	2041	tcagagagccatcagaaagctcatgatgaacaagcttaagaaaaatttagaaggaggtccgaag	2100
Dp	7866	TCAGAGAGCCATCAAGAGCTCATATAGCAAGCTTATAGAAAGGTATGAGAGGGCTCGAAGG	7925
Qy	2101	gaagaagagctctgacccaggggtgtgttgaagaatgtgttcaaacaggtctccctgtgaatgac	2160
Dp	7926	GAAAGAGAGCTGACCAAGGGGTGTTTGAAGATGTGTTCAACAGGTCCTTGCATGTACAC	7985
Qy	2161	acctgtctctctgtctgacggggccccctagctatgctgtccgttactacatgtgg	2220
Dp	7986	ACCCCTGCTTTTCTGCTCTGACGGGACCCCTTATGATGCTCTCTCTGTATTTACAATTTGGG	8045
Qy	2221	acctgtctaactaaatagatgtatgtccctgtgttagaagaacagtgatgtcaatccagaatc	2280
Dp	8046	CTTCTCTTAATTAATAGTGTGTGTGCTTGTGTATGAGAAAGAGTAGTGCAATCCAGATC	8105
Qy	2281	atggtactaagcaacagtatcaaaagcctctgagccaaagagaacatgaccttaagcct	2340

Db	8106	ATGTACTTAGGCAACAGTACCAAGCGCTTTGTGAGCCAAAGAGAAACTGACCTTAGCCT	8165
Qy	2341	tcaccagttcctaagatlagactattacaagaacagaagtgtgggaatgaaagatgaaaa	2400
Db	8166	TCCCAAGTTCAATATTAGAACTATTAAACAGCAAGAAGTGGGAATGAAGAGATGAAAA	8225
Qy	2401	tgcaacctaacccctcccaagaaccagaagttaataaaagctctaaatgccccgaatt	2460
Db	8226	TGCACCTTAACCCCTCCAGAACCCCGAGAAAGTTAATAAAAAAGCTCAATATCCCCCAATT	8285
Qy	2461	acagaaccctgtcgtgcgtccagtaaatagttagaaggttcaacattccattgtgtccagggc	2520
Db	8286	CCAAACCCTCTGGCTGGCCAGTAATAAGTGTGAAGGTGCACCTTCCTAATTGTTCCAGGGCC	8345
Qy	2521	ctgcctatccctggcctaagaataatacaacggaaatgaagttagctctaattgcctatcctgaatt	2580
Db	8346	CTGCTATCCCTGGCGCTTAAGTAACATTAACAGAAATAGTTGACTAATGCTTAATCTGGATT	8405
Qy	2581	ctgtaaacactgactgtgcacatagagaagaattatcaacatttgcacgacctagtgaccta	2640
Db	8406	CTGTAAACAGCACTGGCCACCATTAAGAGATTGATTACATTGCACAGCCCTAAGTGACCTA	8465
Qy	2641	tcttaactgtcaactcttcaactctgtcccagaagcccacagcaatlgcgaacctccgagacta	2700
Db	8466	TCTTAACGTGAATCTGTCACTCTGCCCAAGAACCCACCAATGCGGAGCTCCGGAGCTTA	8525
Qy	2701	ttttaaaatgatttggltccacaagagacgagctcctcgatattttlaaaatgatttggltcca	2758
Db	8526	TTTTAAATATATTGTGTCCACGAGAGCCCGGGCTCTGATATTTTAAATMGATGGTCCA	8583

RESULT	5	9
PEN293657		
LOCUS		
DEFINITION	PEN293657 8763 bp DNA	VRL 30-MAY-2001
ACCESSION	porcine endogenous type C retrovirus proviral gag gene, pol gene	
VERSION	and env gene, class B, clone 213.	
KEYWORDS	AJ293657 GI:14275641	
SOURCE	env gene; envelope; gag gene; group specific antigen; pol gene; polymerase; polyprotein.	
ORGANISM	porcine endogenous type C retrovirus.	
REFERENCE	porcine endogenous type C retrovirus.	
AUTHORS	viruses: Retroid viruses; Retroviridae; Gammaretrovirus.	
TITLE	1 (bases 1 to 8763)	
	Krach, U., Fischer, N., Cauderna, F. and Toenjes, R. R.	
	Comparison of replication-competent molecular clones of porcine	
	endogenous retrovirus class a and class b derived from pig and	
	human cells	
JOURNAL	J. Virol. 75 (12), 5465-5472 (2001)	
MEDLINE	21256017	
REFERENCE	2 (bases 1 to 8763)	
AUTHORS	Toenjes, R. R.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JUL-2000) Toenjes R. R., Medical Biotechnology,	
	Paul-Ehrlich-Institut, Paul-Ehrlich-Str. 51-95, Langen, 63225,	
	GERMANY	

```

LRR
misc_feature
primer_bind
gene
CDS
source
location/Qualifiers
1..8763
/organism="porcine endogenous type C retrovirus"
/proviral
/specific_host="Sus scrofa"
/db_xref="taxon:70540"
/clone="213"
/cell_line="PK15"
/note="class B"
1..630
467
/note="Cap site"
633..650
1077..12651
/gene="gag"
1077..12651
/gene="gag"

```


Db 6769 TCCACCTTTAAACCATTCGCTTTAGGATAGACAGCGGAGACAGACCCTCTGTGGCAGTGGGA 6828
 QY 1081 cccgataaagtactgctgtaacaggggccccggcccttgagccaccgataacttggccg 1140
 Db 6829 CCCGATAAAGTACTGGCTGAACAGGGGCCCCCGCCCTGGAGCCACCGCTAACTTGGCCG 6888
 QY 1141 gtgcccacaaatcaactgcgcgcgcctgacataacacagccgcctagcaacataact 1200
 Db 6889 GTGCCCAATTAACCTGGCTGCGGCGCTGACATACACAGCGCGCTAGCAGACATACCACT 6948
 QY 1201 ggalatgctccacaaacagcgcctagaaactccccagggtgtccctglttaagacagag 1260
 Db 6949 GGATTTGATTCCTACCAACAGCGCTAGAAACTCCCAAGTGTCTCTGTTAAGACAGGACAG 7008
 QY 1261 agactcttcagctcaccagggaggtctcccaagccaacacccacccgacctgagtc 1320
 Db 7009 AGACTCTTCAGTCTCATCCAGGGAGCTTCCAGCCATCACTCCAGCCGACCTGATGCC 7068
 QY 1321 acctctctctgtgctgtgtctctacccagggccctctatlatagaggagatgctaa 1380
 Db 7069 ACTTCTTCTGTGTGGCTTTGTCTATCTCCAGGCGCTCTCTATTATAGGGGATGGCTAA 7128
 QY 1381 gaaagaaatcaatgtgacaaagagacataaataatgatacagtgagggtcccgaaat 1440
 Db 7129 GAAGGAAATTCATGTGACCAAGACATAGAAATCAATGATGCGGGTCCCGAAAT 7188
 QY 1441 aagttacccctcactgaagtccggaggagggagacatgacataagaaagctcccccac 1500
 Db 7189 AAGCTTACCTCTGAGTGAAGTTCGGGAGAGGACATGCTAGGAAAGCTCCGCCATCC 7248
 QY 1501 caccacacacctctgctactgactgtgtgtatagcaggccctcagaanaalcagatatta 1560
 Db 7249 CACCAACACCTTTCATATGACTGTGATTATAGAGCAGCGCTCAGAAATCAGTATTTA 7308
 QY 1561 gtactggttataacaggtgtgtgtgacatgacatctgtgttaacccctgtgtttccacc 1620
 Db 7309 GTACCTGGTTATTAACAGGTGGTGGGACATGCAATGCTGGTTAACCCCTGTGTTCCACC 7368
 QY 1621 tcagctctcaaccaatccaagaattgtgtgtcaatgtgtccaaatcgtcccccgagtac 1680
 Db 7369 TCAGTCTTCAACCAATCCAAAGATTCTGTCTCATGTGTCCAATCGTCCCGAGGTATAC 7428
 QY 1681 taccatcctgagaaggtgtgtctgtatgataatgactacatcgylataacccgacaaaaga 1740
 Db 7429 TACCATCTCGAGGAAGTGTCTGTGATGATATGACTATCGTATTAACCCAGCAAAAGA 7488
 QY 1741 gaaccgctatcccttaccctagctgtgtaatgtcgtataggaagcgcctgtgtgtgga 1800
 Db 7489 GAACCCGTATCCCTTACCTTACCTGTATGCTGTGATTAAGGAGCGCGCTGGGCTGAGGA 7548
 QY 1801 acaaggacagctgcctctgatacacaggacacagcagctagagaagaagactgtgagcta 1860
 Db 7549 ACAGGGACAGCTGCCCTGATATACAGACACACAGCCTAGAGAAAGCACTTGGTGACCTA 7608
 QY 1861 catgcygcacatgacaagaagatcctcgagccttaagaaggtctgttagcaacttaagaag 1920
 Db 7609 CATGGCGCATGACAGAGAGATCTCGAGCCTTACAGCAGTCTGTAGCAACCTGTAGAGAG 7668
 QY 1921 tcccgactcttctgtcgaaggtgtgtctacagaacccgaggaggtatgactctcttt 1980
 Db 7669 TCCCTGACTCTTCTGTCTGAAGTGTGTCTACAGAACCGAGGGGATTAAGATCTCTGTTT 7728
 QY 1981 ctaagaagaagtgtgtctgtcagccttaagaagaagaatgtctctctacgtagatcac 2040
 Db 7729 CTAAGAAGAAGTGTGTGTGTCAGACCTTAAGAAGAAGATGTGCTTGTATGATGATCAC 7788
 QY 2041 tcaaggacacatcagaagactcagacaagaagcttagaanaaaglttagaaggcgtcgaaag 2100
 Db 7789 TCAGAGAGCATCAGAGACTCCATGAGCAAGCTTACAGAAAGATTAGAGAGCGCTCGAAGG 7848
 QY 2101 gaaagaagagctgaccaggggtgtgttgaagatgtgtcacaaggtcctctgatatgcc 2160
 Db 7849 GAAAGAGAGGCTGACCCAGGGGTGTGTGAAGATGTTCACAGGTCTCTCTGTGGATGACC 7908

QY 2161 accctgcttcctcctgacggggccctagtagtccctgtctcctgttactaacgttgg 2220
 Db 7909 ACCCTGCTTTCGCTCTGACGGGGCCCGCTAGTAGCTGCTCTGCTTACTTAACATGTGG 7968
 QY 2221 ccttgcttaataataggtgtgtccttctgttagagaagagtgatgacagatc 2280
 Db 7969 CTTGCTTAATTAATAGGCTTGTGCTTGTGTAAGAACGAGTGTGATCCATCCAGATC 8028
 QY 2281 atgtagtacttaggcaagaagraccaagcctctcagccaagggagaactgactagcct 2340
 Db 8029 ATGCTACTTAGGCAACAGTACCAAGGCTTCTGAGCCAGAGGAAACTGACTTACGCT 8088
 QY 2341 tcccaatctaaagatagacatctaaacagaacaaagtgaggagtgaaagatgaaaa 2400
 Db 8089 TCCCACTTCTTAAGATTAGAACTATTACAGAACAGAAAGTGGGATGAAGATGAAAA 8148
 QY 2401 tgcacactaacctccccaagaccagaagaaatgataaagaactctaaatgccccgaat 2460
 Db 8149 TGCACACTTAACCTCCCAACACCCAGAGAGTTAAATAAAAAGCTTAAATGCCCCGAAT 8208
 QY 2461 acagaccctgctgctgctgcagtaaataggtagaaggtcaactcctatgttccaggac 2520
 Db 8209 CCAAGACCTGCTGCTGCTGCCAGTAATAGTACAGAGTCACTTGTCTATTGTCCAGGC 8268
 QY 2521 ctgctactcctgctcctaaagtaacagagaagaaatgagtgactaatcgctatctggat 2580
 Db 8269 CTGCTATCTCGGCTTAAGTAAGTACAGAAATGAGTTAGACTTAATCGCTATCTGAGAT 8328
 QY 2581 cgttaaaactgactgacacacataagaagaattgattacacatgagaagcctatgacta 2640
 Db 8329 CTGTAAGCTGACTGCGACCATTAAGAAATGATTACATGATGACAGCCCTAGTGACTA 8388
 QY 2641 tct 2643
 Db 8389 TCT 8391

RESULT 10
 AF014162
 LOCUS AF014162 1971 bp mRNA VRL 01-JUN-1998
 DEFINITION porcine endogenous retrovirus envelope protein (env) mRNA, complete cds.
 ACCESSION AF014162
 VERSION AF014162.1 GI:3169735
 KEYWORDS
 SOURCE
 ORGANISM
 porcine endogenous retrovirus.
 Viruses; Retroviral viruses; Retroviridae; Mammalian type C
 retroviruses; 1-Mammalian type C virus group.
 REFERENCE
 1 (bases 1 to 1971)
 Haworth, C., Galbraith, D. N., Lees, G. M. and Smith, K. T.
 Porcine endogenous retrovirus (PoEV) Env sequence
 unpublished
 2 (bases 1 to 1971)
 Haworth, C., Galbraith, D. N., Lees, G. M. and Smith, K. T.
 Direct Submission
 Submitted (16-JUL-1997) Q-One Biotech Ltd., Todd Campus, Acre Rd.,
 Glasgow G20 0XA, UK
 FEATUERS
 source
 1. 1971
 /organism="porcine endogenous retrovirus"
 /db_xref="taxon:61673"
 /cell_line="derived from cell line PK15"
 /note="PoEV"
 1. 1971
 /gene="env"
 1. 1971
 /gene="env"
 1. 1971
 /codon_start=1
 /product="envelope protein"
 /protein_id="AAC17948.1"
 /db_xref="GI:3169736"

/translation="MHPTLSRHLPTRGEPKRIPLPLSFASIAWMLTLTITPQASSK
RLIDSSNHPRLSLJLWLIIDPDGTIVNSTRGAWAPRCITWPELHFCILRLINPVSFNP
PLIVRSYFCCPCTEKREKYGSGSESVCSNRSGVSTNDMDKMPISLJORVAFSPFN
SGPKYKMKMLYKDKSCSPDLIDYKLSIFERKTKSKYDKMLKELGNSPLILYGGAG
STFLIRLRIENGTEKVPKAMPDYLKBOGPALPEPNLIPVPLTSLRPIITPPNS
TTGLIPTNPNRNSGVYKQORLFSILQAFQAINSTDADTSSCLCISGPPYE
GMAKERFNVYKERNOCITWGRNKLTLTVSGDCTGAPSHHLCYSTVYEDQ
SENYLIVPGYNRMWACNTGLPCVSTSVFNOGDFCVMQIVPRVYHHPREVLDEYD
YRNRPKEVPSLTLAVMLGLTAVMLGTALINGPOOLEGKGLBLHAMEDLRA
LKEVSINLESLTSLSEVLONRGLDLFLREGGLCAALKEGCFYDHSGLIRDSM
NKLKRLERRREREADOGWPEGNFNSPMPTLISLTPPLVLLLLLTGVGCLINR
FVAFVRRVRSVAVQIMVLRQOYGLISGEFTDL"

BASE COUNT 530 a 490 c 482 g 469 t
ORIGIN

Query Match 50.7%; Score 1499; DB 14; Length 1971;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 383 ggccgcaactccgactcggggtgagagccgaaagactagatcccttaagcttcg 442
Db 20 GGGCCGACCTCCGACCTGGGGTGAAGCGGAAAGCTGAGATCCCTTAAGCTTCG 79
QY 443 cctccatcgctggtccttaactaacaactcccgagccagtagtaaacgctta 502
Db 80 CCTCCATCGCTGGTTCCTTACTTAACTAACATTAACCTCCGAGCCAGTAGTAACGCTTA 139
QY 503 tagaagctggaaccccatagaccttatccctactgctgctatattgacctata 562
Db 140 TAGACAGCTGACACCCCATAGACCTTATCCCTGAGCTGATTAATGACCTGATA 199
QY 563 cgggtgtcactgtaaatagacactcgaaggtgtgctccttagagagcctgtgagcctaac 622
Db 200 CGGCTGTCACGTGAATAGACACTGAGAGTGTGCTCTAGAGGACCTGGTGGGCTGAC 259
QY 623 tgccttctgctcctcgatgatttaaccccgctgttaaaagacacactcccaactagctcc 682
Db 260 TGCAATTCCTGCTCCGATGATTAACCCCGCTGTAAAGCACACCTCCACCTACTATGCC 319
QY 683 gtaattatggttctatctgctcgcagacagagaaagaaatactgtgagaggtcttg 742
Db 320 GTACTTATGGGTTTATATTGCTGCCGACGACAGAGAAAGAAATACTGTGGGGTCTTG 379
QY 743 gggaaatccctctgtagagagtgaagctgcgtcaactccaaagctggaagctggaagctgc 802
Db 380 GGGAAATCCTTCTGTAGAGAGTGAAGCTGCTCACTCAACGATGGAAGTGGAAATGCC 439
QY 803 ggaactctccagagacgggtaaaatctctcttctgcaattccggccgggcaagtata 862
Db 440 GGAATCTCTCCAGAGCGGGTAAATTCCTTGTCAATTCGCGCCGGGCAAGTACA 499
QY 863 aaatgtagaaactataaaagataagagctgctcccatcagactagactatcctaaga 922
Db 500 AAATGATGAACATATATAAGATAGAGCTGCTCCCATAGACTTATGATTAATTAAGA 559
QY 923 taagttcaactgaaagaaagaaataatcaaaagtggataaagtgtatagct 982
Db 560 TAAATTTCACTG--AAAGGAAACAGGAAATATTCAAAAGTGAATGAATGATGAGCT 617
QY 983 ggggaagaagtgtttttaaataatgaggggagcagaggtccacttaacactcgcctta 1042
Db 618 GGGGAATAGTTTTTT-ATTATATGGCGGGGAGGAGCTCCACTTAAACCATTCGCGCTTA 676
QY 1043 ggaatagagacgggagacaaacccctgtgcaatggagaccggaataaagtactgctgaac 1102
Db 677 GGAATAGAGAGCGGGAGACAAACCCCTGTGCAATGGAGACCCGATTAAGTACTGGCTGAAC 736
QY 1103 aagggcccccggccctgagagcaacgcaataactgcggtgtgcccataaactcgtcgc 1162
Db 737 AGGGGCCCCCGCCCTGAGACCCGCAATTAATTCGCGGTCGCCCAATTAACCTCGCTGC 796
QY 1163 ggcctgaacataacagccgcttagcaaacgtaacactgattgattcctcaacaacagc 1222

Db 797 GGCTGACATTAACACACCCGCTAGCAACAGTACCACTGATGATGATTCACCAACACCG 856
QY 1223 ctgaaacactcccaagtggtccgtgtaagaagcagagagactcctcagctcaatccag 1282
Db 857 CTGAAACCTCCCAAGGTTGCTGTTAAAGACAGAGACAGACCTTCAATCTCATCAAG 916
QY 1283 gaacttccaagcaactcaactcaccgagaccctgaatgcactctctctgttgactgtgc 1342
Db 917 GAGCTTCCCAAGCATCACTCACTCAGGACCCCTGATATCCACTTCTTGTGCTTGTGTC 976
QY 1343 tatcctcagggcctccttattatgagggatggctaaagaagaataatcaatgtgaca 1402
Db 977 TATCTCAGAGGCTCCTTATATGAGGGATGCTAAAGAAAGAAATCAATGTGACCA 1036
QY 1403 aagagatgaataatcaatgatactgaggggtcccgaaataagcttaaccccaagatt 1462
Db 1037 AAGAGCATGAATAATCAATGATGATGAGGGGCTCCGAAATTAAGCTTACCTCACTGAAGTTT 1096
QY 1463 ccggaagggagacatgataagaaagctcccccatacccaacaacacttgcctatagta 1522
Db 1097 CCGGGAAGGGGACATGATGAGAAAGCTCCCATCCACCAACACTTGTCTTATAGTA 1156
QY 1523 ctgtgttatagagcagggcctcagaataatcaglatatgtaacctgttataaaggtgtc 1582
Db 1157 CTGTGTTATATGAGCAGGCTCAGAAATCAGATTAATTAACCTGTTATTAACAGTGTG 1216
QY 1583 gggcagcaaatcactggtgttaacccctgtgttccactcagctctcaacaatccaaag 1642
Db 1217 GGGCATGCAATATCTGGGTTAACCCCTGTGTTCCACCTCACTTCAACCAATCCAAAG 1276
QY 1643 atttgtgtcactggtcccaatcgtcccccaggtgtactacacactcagaaggtgtcc 1702
Db 1277 ATTTCGTGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1336
QY 1703 ttgaatgaatgactatcggtlaaacgagccaaagagagacglatcccttaacctag 1762
Db 1337 TTGATGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1396
QY 1763 ctgttaactgctgattagagagcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1822
Db 1397 CTGTAATGCTCGGATTAAGGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1456
QY 1823 caggaacacagagcttagaagaagactgtgtgactacatcagccatcagaaagatc 1882
Db 1457 GAGGACACAGCAGCAGTGAAGAAAGACTTGTGTGACTTATGATGATGATGATGATGATG 1516
QY 1883 tccgagccttaagaaggtcgtttagcaactagaaagctccctgactctctgtcgaag 1942
Db 1517 TCCGAGCCTTAAAGGAGTCTGTAGCAACCTGGAAGAGTCCCTGACTTCTTGTGTGAAG 1576
QY 1943 tggttctacagagccggaagggatagatctgtctgttcttaagaagaaggtgtatgt 2002
Db 1577 TGGTTCTACAGAACCGAGGGGATTAAGATCTGCTCTTCTTAAGAGAAAGGGGTTATGTG 1636
QY 2003 cagccttaaaagaagaatgtgtcttctatgtatcaactcagagagcactca 2062
Db 1637 CAGCCTTAAGAAAGAAAGTGTCTCTTAATGATGATGATGATGATGATGATGATGATG 1696
QY 2063 tgaacaagcttaagaagaagaatgagagagcgtcgaagggaagagagagctgaaccaaggtc 2122
Db 1697 TGAACAAGCTTGAAGAAAGAAAGTGAAGAGCGTCGAGAGGAAAGAGAGGCTGACACGAGGCT 1756
QY 2123 ggttgaagaagatgtgtcaacaaggtcctcttgatgagacacactgctctctcgaag 2182
Db 1757 GGTGTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1816
QY 2183 ggcctcagtagtccctgtcctgttacttaagtagtgggctgtgcttaataatagtttg 2242
Db 1817 GGCCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1876
QY 2243 ttgccttgtttagaagaagagtagtgcagtcacagatcaatgttacttaggcaacagtagc 2302

Db	1877	TTGCCCTTTGTTAGACAGAGTGAGTGCAGTCCAGATCAATGCTACTTAGGACCAAGTACC	1936
QY	2303	aagccctctgagcccaagagaactgacctag	2337
Db	1937	AAGGCTTCTGAGCCAGGAAACTGACTCTAG	1971
RESULT	11		
AF038601			
LOCUS	AF038601	7333 bp	MRNA
DEFINITION	Sus scrofa porcine endogenous retrovirus ERV-PK15 mRNA, complete		
ACCESSION	AF038601		
VERSION	AF038601.1	GI:3133304	
KEYWORDS			
SOURCE			
ORGANISM	Sus scrofa		
REFERENCE	Akiyoshi, D.E., Denaro, M., Zhu, H., Greenstein, J.L., Banerjee, P. and Fishman, J.A.		
AUTHORS			
TITLE	Identification of a full-length cDNA for an endogenous retrovirus of miniature swine		
JOURNAL	J. Virol. 72 (5), 4503-4507 (1998)		
MEDLINE	98216827		
REFERENCE	2 (bases 1 to 7333)		
AUTHORS	Fishman, J.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-DEC-1997) Infectious Disease Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA		
REFERENCE	3 (bases 1 to 7333)		
AUTHORS	Fishman, J.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAY-1998) Infectious Disease Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA		
REFERENCE	Sequence update by submitter		
REMARK	On May 16, 1998 this sequence version replaced gi:3116445.		
COMMENT	Location/Qualifiers		
FEATURES	1..7333		
source	/organism="Sus scrofa"		
	/strain="miniature swine"		
	/db_xref="taxon:9823"		
	/cell_line="PK-15"		
	/note="porcine endogenous retrovirus ERV-PK15; infective virus but represents a naturally occurring transcript"		
LTR	1..578		
CDS	598..2172		
	/codon_start=1		
	/product="gag protein"		
	/protein_id="AAC16766.1"		
	/db_xref="GI:3116446"		
	/translation="MGQTVTPPLSLTLDHWTNRSRAHNLSVQYKKGHWQTFCASEMP TFDVGWPESTGFNSLILAVKAILFQTPGSHPOEPYIILWDLADBPWPVWPMN KPKRGPRIILALGKRNKSAKVEPSPIYPIEIEPTWPPVPPYPAQCAVAG PSAPGAVVVGPAAGTGRSRGATPERDEIAILPLRYGPPMGGOLOPVOYFSS ADLVNWKTHPEPSEDPORLTGLVLSLMSHQPTMDCCOOLLOLTLTTEERILLPA KKNVGAGRPTOLONTDMGFPPLTRBGMWNTNAGRSKATLYQALVAGIRGASRP TNLAKVREVMGPNPPSVFLERLMEARRTTPDPITSEAKASVALAFIGSALDIR KILQLEGLDAELRLDLVREAEKYVYRETEEEKEQREKREERERDRORCKNT KILAAVVEGKSRERERDFRKIRSGPROSGNLGNRTPLDKQCAVCEKKGHWANCPK KGNKGPVLALEDKD"		
	2320..3552		
	/note="truncated protein"		
	/codon_start=1		
	/product="pol protein"		
	/protein_id="AAC16767.1"		
	/db_xref="GI:3133305"		
	/translation="MGATGROYPMTRRTVDLGVHVSFLVTPCVPILGRDL TKGAQISIEGGRPEVSNKPLITVLQDLDEERYLSPOVKPRQDQSNIEQPPQW AETAGMGLAKVPPVOYLOKASATPVSRYOTPLSREREGIMHVOQLDQGLTLPVQ SPWNPPLVPRKPGTNDIRPVQDLREVKRQVDLHPTVPEYNNLSALPPERMVTVL DLKDAFELRLHPTNSQPLTFPEMRDPGIRGQGLTWTRLPDGFKNSTPIDEALHRLD		
CDS			

Query Match	27.5%	Score 813;	DB 4;	Length 7333;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches	863;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
ANRRIQHPQVTLQYDDLLLAGATKODCLEGTAKALLLESLDIGRASAKKAOICRE VLYLGSYLRSRGORMLTEARKRTVVOIIPAPTTARQVREFISTAGFCRLMIPGATLAP LVLPLTRKGGCCLPQGGSK"				
4803..6725				
/note="env protein; N terminus deleted"				
/codon_start=1				
/pseudo				
6726..7333				
BASE COUNT	1984 a	1788 c	1885 g	1676 t
ORIGIN				

QY	2082	glttagagagcgctcgaaggaagagaagagctgaccaggggtgtgttaaggaagtgtcaa	2141
Db	6470	GTTAGAGAGGCGCTGAAAGAGAAAGAGAGGCTGACAGGGGTGTGAAAGATGTTCAA	6529
QY	2142	caggtctctctggaatgaccaccctgtcttctgctctgaaggggccccctagtgcctgct	2201
Db	6530	CAGGCTCTCTTGATGATGACCACTGCTTCTTCTGCTCTGACGGGGCCCTAGTAGTCTGCT	6589
QY	2202	cctgttacttacaagltgggacctgtcttaataataagtttltgaccttgttagaagaq	2261
Db	6590	CGTGTTACTTACAGTTGGGCGCTTGCTTAAATTAATAGTTGTGCTCCPTTGTAAAGAACG	6649
QY	2262	agtgagtgcaatccaagatcatgttacttaggcaacaagtaacaaagcctctgagccaag	2321
Db	6650	AGTGAGTCACTCCAGATCATGACTTAGGCAACAGTACCAAGGCGCTTCTGACCCAAAG	6709
QY	2322	agaactgaccttgaagcttccagcttctgaatgaatgaatgaatgaatgaatgaatgaat	2381
Db	6710	AGAACTGACCTCTTACCTTCCAGTTCTTAGATTAGACTATTATTAACAAGCAAGAGTG	6769
QY	2382	gggaatgaaagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa	2441
Db	6770	GGAATGAAAGATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA	6829
QY	2442	ctctaataatgccccgaatattacagaccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2501
Db	6830	CTCTAAATGCCCCCGGAATTCAGACCCGCTGCTGCGCAGTAATAGTGAAGAGTGCACA	6889
QY	2502	cttctatgttccaagggccctgtatctctgtgcttcaatgaatgaatgaatgaatgaatga	2561
Db	6890	CTTCTTATTTGTTCCAGGGCGCTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	6949
QY	2562	ctaactgcttacttgatctgtlaaaactgactgacacatagagaatgatatcat	2621
Db	6950	CTAATCGCTTATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	7009
QY	2622	tgaagaccctagtgacctatctcaactgcaactgtgtcaactgtgtgtgtgtgtgtgtgtgt	2681
Db	7010	TGACAGCCCTAGTACCTATCTCAACTGCAATGTGTCACCTGTGCCAGGACCCACGACG	7069
QY	2682	atggagacctcggagactatttaaatgatatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2741
Db	7070	ATGGGACCTCTCGGAGCTATTTAATGATTGTGTCACGAGCGCGGCTCTCGAATAT	7129
QY	2742	ttaaatgatatgt	2801
Db	7130	TTAAATGATTTGCTGCATGAGCGCGGCTCTGATATTTTAAATGATTTGCTGTGTGAC	7189
QY	2802	gcacagccttgt	2861
Db	7190	GCAAGAGCTTGTGTGTGAACCCCATTAAGACTGCGGATTCGCACTCGGAGCCGCGAGT	7249
QY	2862	ccttacccttgcgt	2921
Db	7250	CTCTTACCCCTGCGT	7309
QY	2922	tgctgttgcatacaaaaaaaaaa	2945


```

FT      /*tag= f
FT      /product= ENV protein
FT      /note= "envelope protein"
FT      LTR
FT      7591..8209
FT      /*tag= g
FT      7638..8106
FT      /*tag= h
FT      /note= "U3 LTR domain"
FT      7643..7686
FT      /*tag= i
FT      /note= "Binding site for transcription factors
FT      OCF-1, c-Myb, LfF-1, E47 and ETS-1"
FT      7719..7739
FT      /*tag= j
FT      /note= "Binding site for transcription factors
FT      AP-4 and NF1"
FT      7744..7754
FT      /*tag= k
FT      /note= "Binding site for transcription factor AP-1/TR"
FT      7773..7785
FT      /*tag= l
FT      /note= "Binding site for transcription factor ETS-1/GATA"
FT      7793..7834
FT      /*tag= m
FT      /note= "Binding site for transcription factors
FT      GATA, ETS-1, c-Myb, AP-1, GATA, E47"
FT      7841..7851
FT      /*tag= n
FT      /note= "Binding site for transcription factor AP-1"
FT      7873..7883
FT      /*tag= o
FT      /note= "Binding site for transcription factor AP-1"
FT      7885..7914
FT      /*tag= p
FT      /note= "Binding site for transcription factors
FT      AP-1/GATA, c-Myb and AP-1"
FT      7928..7953
FT      /*tag= q
FT      /note= "Binding site for transcription factors E47
FT      and ETS-1"
FT      7964..7968
FT      /*tag= r
FT      7985..7997
FT      /*tag= s
FT      /note= "Binding site for transcription factor GATA"
FT      8003..8007
FT      /*tag= t
FT      8024..8036
FT      /*tag= u
FT      /note= "Binding site for transcription factor GATA"
FT      8042..8046
FT      /*tag= v
FT      8049..8060
FT      /*tag= w
FT      /note= "Binding site for transcription factor
FT      AP-1/CRAB"
FT      8075..8090
FT      /*tag= x
FT      1..61
FT      /*tag= y
FT      /note= "R LTR domain"
FT      8127..8135
FT      /*tag= z
FT      /note= "PolyA downstream element"
FT      8161..8166
FT      /*tag= aa

```

```

PR      10-FEB-1997; 97GB-0002668.
PR      19-APR-1996; 96GB-0008164.
XX
PA      (TMUT-) IMUTRAN LTD.
PA      (QONE-) Q-ONE BIOTECH LTD.
XX
PI      Galbraith DN, Haworth C, Lees GM, Smith KT;
DR      WPI: 1997-535851/49.
XX
P-PSDB; AAM39271; AAM39272; AAM39273.
XX
PT      Polynucleotide encoding porcine retrovirus expression product -
XX      useful to develop products for use in vaccines, diagnosis and
XX      xeno-transplantation
XX
PS      Claim 4; Fig 3; 69pp; English.
XX
CC      This DNA sequence encodes the porcine retrovirus (POEV) virion core
CC      polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins and
CC      also includes the long Terminal Repeat (LTR). These proteins can be used
CC      to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC      antiviral agents. They can also be used in xeno-transplantation
CC      technology and as diagnostic tools.
XX
SQ      Sequence 8209 BP; 2168 A; 2064 C; 2154 G; 1823 T; 0 other;

```

```

Query Match      100.0%; Score 8209; DB 18; Length 8209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 gtagtgtaagctgtggcccgccgagcgttggaataaataatcctctgtgttcac 60
DB      1 gtagtgtaagctgtggcccgccgagcgttggaataaataatcctctgtgttcac 60
QY      61 aagacgcttcctgtgagtgatgtgggtgtgtcctcttcgagccgagcagagggagat 120
DB      61 aagacgcttcctgtgagtgatgtgggtgtgtcctcttcgagccgagcagagggagat 120
QY      121 tgtcttcttaagcgttccttcatttggtgtgtgtgcccggaaatcctgtgacacccctta 180
DB      121 tgtcttcttaagcgttccttcatttggtgtgtgtgcccggaaatcctgtgacacccctta 180
QY      181 caaccgagaaacgacttgagtgagtgaaagatccctcttgaaagcgtgtgtgtcggcc 240
DB      181 caaccgagaaacgacttgagtgagtgaaagatccctcttgaaagcgtgtgtgtcggcc 240
QY      241 ggcgtctcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 300
DB      241 ggcgtctcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 300
QY      301 ctcagacgtaagcgttgagcgttgatcagcagcagcgttgagagagacagagctg 360
DB      301 ctcagacgtaagcgttgagcgttgatcagcagcagcgttgagagagacagagctg 360
QY      361 ccaccctggggagacgcccggaggttgaggagacgagcagcgttggttccctact 420
DB      361 ccaccctggggagacgcccggaggttgaggagacgagcagcgttggttccctact 420
QY      421 gtcggtcagagagacgagctcgttctgttgaaagcgaagcctccctcgcgcgtccg 480
DB      421 gtcggtcagagagacgagctcgttctgttgaaagcgaagcctccctcgcgcgtccg 480
QY      481 actctttgctcgtcgttgaggagagcagcagcgtcgtgtgtcgtgattcgtgtgtt 540
DB      481 actctttgctcgtcgttgaggagagcagcagcgtcgtgtgtcgtgattcgtgtgtt 540
QY      541 ctgtctcgtgtgtcttctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 600
DB      541 ctgtctcgtgtgtcttctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 600
QY      601 tgactaccccttagtttgactcgcacattgactgaagttagatcagaggtcata 660
DB      601 tgactaccccttagtttgactcgcacattgactgaagttagatcagaggtcata 660

```

Db 601 tgactcccccttagttgactctcgacatctgactgaatgaatgataccagggctcata 660
Qy 661 attgtcagttcaaggttaagaaagaccttgcagactttctgtcctcgaaatggccaa 720
Qy 661 attgtcagttcaaggttaagaaagaccttgcagactttctgtcctcgaaatggccaa 720
Qy 721 cattgatgttgatgacatcagaggggacatttaattctgaaatatactctgacttga 780
Db 721 cattgatgttgatgacatcagaggggacatttaattctgaaatatactctgacttga 780
Qy 781 aggaacatcatttccagacttgaccccggtctcatccgatacgaagccataatccctta 840
Db 781 aggaacatcatttccagacttgaccccggtctcatccgatacgaagccataatccctta 840
Qy 841 cgtgagcaagatttgagcaagatctcccgcaatgggttaaacatctgcttaataaccaa 900
Db 841 cgtgagcaagatttgagcaagatctcccgcaatgggttaaacatctgcttaataaccaa 900
Qy 901 gaaagccaggtcccggaatccctggtctcttgagagaaagaaacacatccgagcaaaag 960
Db 901 gaaagccaggtcccggaatccctggtctcttgagagaaagaaacacatccgagcaaaag 960
Qy 961 tcgagccctctctctgatactaccccgagatcgagagagccgagacttgcggaaccc 1020
Db 961 tcgagccctctctctgatactaccccgagatcgagagagccgagacttgcggaaccc 1020
Qy 1021 caacgtgttcccccaccccttaccagacaaaggtgtgttgagagagacactctgcccctc 1080
Db 1021 caacgtgttcccccaccccttaccagacaaaggtgtgttgagagagacactctgcccctc 1080
Qy 1081 ctgagagctccggttgagagagacactctgagcgagactcgagagcgagagagcgcaacc 1140
Db 1081 ctgagagctccggttgagagagacactctgagcgagactcgagagcgagagagcgcaacc 1140
Qy 1141 cggagcgagacagagagatcgagatataccggtcgagacatacggccctcccatcgcaag 1200
Db 1141 cggagcgagacagagagatcgagatatacgggtcgagacatacggccctcccatcgcaag 1200
Qy 1201 ggggccaattggagccctccagatattgaccccttctctctgagatctataatggga 1260
Db 1201 ggggccaattggagccctccagatattgaccccttctctctgagatctataatggga 1260
Qy 1261 aaaaataacatccccccttctcgagagatcccccagcgctcaggggttgagagttccc 1320
Db 1261 aaaaataacatccccccttctcgagagatcccccagcgctcaggggttgagagttccc 1320
Qy 1321 ttatgttctctcaccagacttacttgagatgtgtcaacagctgtcagacactctca 1380
Db 1321 ttatgttctctcaccagacttacttgagatgtgtgtcaacagctgtcagacactctca 1380
Qy 1381 caacgagagagagagagaaattctgttaagagctagaanaaatctatcgcaagctctgtg 1440
Db 1381 caacgagagagagagagaaattctgttaagagctagaanaaatctatcgcaagctctgtg 1440
Qy 1441 ggcgacccaagcaatgttgaaatctgaatgagatttccctgactcgcccggtt 1500
Db 1441 ggcgacccaagcaatgttgaaatctgaatgagatttccctgactcgcccggtt 1500
Qy 1501 gggactacaacacggtgaaagttagggagagctgaanaaatctatcgcaagctctgtg 1560
Db 1501 gggactacaacacggtgaaagttagggagagctgaanaaatctatcgcaagctctgtg 1560
Qy 1561 cgggtctccggggcgcttaagaagcgccactaatcttgctaagagtaagaaggtgagtc 1620
Db 1561 cgggtctccggggcgcttaagaagcgccactaatcttgctaagagtaagaaggtgagtc 1620
Qy 1621 agggacgaagacacactccctcggtattcttgagagagctatggaagccctcagcggt 1680
Db 1621 agggacgaagacacactccctcggtattcttgagagagctatggaagccctcagcggt 1680
Qy 1681 tcaccccttttgatctacactcagagagccagaaagcctcagtgccctgcttcatgt 1740
Db 1681 tcaccccttttgatctacactcagagagccagaaagcctcagtgccctgcttcatgt 1740

Qy 1741 ggcagtcgctctgatatcagagaaagaaacttcagagacttgaaaggttacaagaggtc 1800
Db 1741 ggcagtcgctctgatatcagagaaagaaacttcagagacttgaaaggttacaagaggtc 1800
Qy 1801 agttacgttgatctagtgagagagcgagagaggtgtatatacgaagagggagacagaaag 1860
Db 1801 agttacgttgatctagtgagagagcgagagaggtgtatatacgaagagggagacagaaag 1860
Qy 1861 agaaagaaagagaaagaaagagagagagaaagagagagagagagagagagagagagag 1920
Db 1861 agaaagaaagagaaagaaag 1920
Qy 1921 aagaaagaaatctgaactaagatcttgcgcaggtgttgaaagagagagagagagagag 1980
Db 1921 aagaaagaaatctgaactaagatcttgcgcaggtgttgaaagagagagagagagagag 1980
Qy 1981 gagagagagatcttaggaanaattaggttcagagccctagaacagtcagagagagagag 2040
Db 1981 gagagagagatcttaggaanaattaggttcagagccctagaacagtcagagagagagag 2040
Qy 2041 ggaacccactcgacaaagagacaggtgtgctatgttaagaaagagacactgggcaagga 2100
Db 2041 ggaacccactcgacaaagagacaggtgtgctatgttaagaaagagacactgggcaagga 2100
Qy 2101 actgcccagaagagagagaaacaaagagacgaaggtctcagacttgaaagagataagatt 2160
Db 2101 actgcccagaagagagagaaacaaagagacgaaggtctcagacttgaaagagataagatt 2160
Qy 2161 agggagagagaggttctcgagccctcccgagcccgagccagagtaacttgaaagtgagggc 2220
Db 2161 agggagagagaggttctcgagccctcccgagcccgagccagagtaacttgaaagtgagggc 2220
Qy 2221 aaccagttgagttctgtgtgataccgagagcgagacattcaggtctgataacacattag 2280
Db 2221 aaccagttgagttctgtgtgataccgagagcgagacattcaggtctgataacacattag 2280
Qy 2281 gaaacctaagaaagaaanaaaatctctgtgtatgggtgccaagagcgaaagcgagatccat 2340
Db 2281 gaaacctaagaaagaaanaaaatctctgtgtatgggtgccaagagcgaaagcgagatccat 2340
Qy 2341 ggaactaccgaaagacggtgacttggagatggagagaggttaacccactcgtctctgta 2400
Db 2341 ggaactaccgaaagacggtgacttggagatggagagaggttaacccactcgtctctgta 2400
Qy 2401 tcccgaggtgcccagtaaccccttctagtgagagacttactgacacagagatggagctcaaa 2460
Db 2401 tcccgaggtgcccagtaaccccttctagtgagagacttactgacacagagatggagctcaaa 2460
Qy 2461 ttctcttgaacaaagaaagacgaagatgtctgtgaaataacaaacacacacacacacac 2520
Db 2461 ttctcttgaacaaagaaagacgaagatgtctgtgaaataacaaacacacacacacacac 2520
Qy 2521 cccctcaatataatgatacgaactatacttcccccagtaaaagcctgataaagata 2580
Db 2521 cccctcaatataatgatacgaactatacttcccccagtaaaagcctgataaagata 2580
Qy 2581 tacagttcgtgtgagagaggttccccaagcctggagcgaaacggcagaggtgtgtg 2640
Db 2581 tacagttcgtgtgagagaggttccccaagcctggagcgaaacggcagaggtgtgtg 2640
Qy 2641 caaagcaagttcccccacaggtatatacactgaagagcgagtgctacacacacacacacac 2700
Db 2641 caaagcaagttcccccacaggtatatacactgaagagcgagtgctacacacacacacacac 2700
Qy 2701 gacagtaaccccttgagtagagaggtccgagagagaaatttgccgcaatgttcaagatcaa 2760
Db 2701 gacagtaaccccttgagtagagaggtccgagagagaaatttgccgcaatgttcaagatcaa 2760
Qy 2761 tccaacagagagatcctgtctcgtcccaatcccttggaatacctccctgtcaccaggtta 2820
Db 2761 tccaacagagagatcctgtctcgtcccaatcccttggaatacctccctgtcaccaggtta 2820

QY 2821 ggaagcctggagccaatgatattatcgaccagctgaagagctgaagaggtcaataaagg 2880
DB 2821 ggaagcctggagccaatgatattatcgaccagctgaagagctgaagaggtcaataaagg 2880
QY 2881 tgcagagacatacaccacaagctcccgaaacctataacctcttgagcgccctccgctg 2940
DB 2881 tgcagagacatacaccacaagctcccgaaacctataacctcttgagcgccctccgctg 2940
QY 2941 aacggaactgttacacagatgttgacttaaaagatgctctctctctgctgattacacc 3000
DB 2941 aacggaactgttacacagatgttgacttaaaagatgctctctctctgctgattacacc 3000
QY 3001 ccaactagccaaccaacttttgccttcgaaatgagagatccaggtacaggaagaaacgggc 3060
DB 3001 ccaactagccaaccaacttttgccttcgaaatgagagatccaggtacaggaagaaacgggc 3060
QY 3061 agctcaacctggagccagcagcccaaggttcaagaactcccgacaacctcttgagagag 3120
DB 3061 agctcaacctggagccagcagcccaaggttcaagaactcccgacaacctcttgagagag 3120
QY 3121 ccttacacagagagccttgccaactcagagatcccaacctcaggtgacccctccaggt 3180
DB 3121 ccttacacagagagccttgccaactcagagatcccaacctcaggtgacccctccaggt 3180
QY 3181 acgtgagatgacctgtcttcgagcgagagccacaacagagactgttagaaggtacagag 3240
DB 3181 acgtgagatgacctgtcttcgagcgagagccacaacagagactgttagaaggtacagag 3240
QY 3241 cactacgtcgtgaattgtctgaacctgaagctacagagcctctgtctgaagagccagatt 3300
DB 3241 cactacgtcgtgaattgtctgaacctgaagctacagagcctctgtctgaagagccagatt 3300
QY 3301 gcaagagagaggtgaacatacttgggttaacagtttgcggcgagagcgatgtgctgagag 3360
DB 3301 gcaagagagaggtgaacatacttgggttaacagtttgcggcgagagcgatgtgctgagag 3360
QY 3361 aggcagaggaagaaactgtatgctacagataccgcccacaacagccagcaaacagtgaag 3420
DB 3361 aggcagaggaagaaactgtatgctacagataccgcccacaacagccagcaaacagtgaag 3420
QY 3421 agttttgggagagctgtgattttcagagctgtgagatccgggttttgagaccttagcag 3480
DB 3421 agttttgggagagctgtgattttcagagctgtgagatccgggttttgagaccttagcag 3480
QY 3481 ccccactctaccgcttaacccaagaaaggaggtcttccctggtcctctgaagcacaga 3540
DB 3481 ccccactctaccgcttaacccaagaaaggaggtcttccctggtcctctgaagcacaga 3540
QY 3541 aggcatttgatgtatcaaaaagccctgtcagcagcctgtcctctgagccctccagag 3600
DB 3541 aggcatttgatgtatcaaaaagccctgtcagcagcctgtcctctgagccctccagag 3600
QY 3601 taactaaaccttaaccttattatgtatgagcgttaagaggttagccgagagagtttaa 3660
DB 3601 taactaaaccttaaccttattatgtatgagcgttaagaggttagccgagagagtttaa 3660
QY 3661 cccaacaccttaagagccttgagagagagcctgtctgaactgttcaagaagagctgtac 3720
DB 3661 cccaacaccttaagagccttgagagagagcctgtctgaactgttcaagaagagctgtac 3720
QY 3721 taagcagttgttgccgtatgtctgaagagctatcgagcgtgtgccaatactgtac 3780
DB 3721 taagcagttgttgccgtatgtctgaagagctatcgagcgtgtgccaatactgtac 3780
QY 3781 agcgtgaacaaatgactttggagagaaatataactgtatagcccccaatgactgtgaga 3840
DB 3781 agcgtgaacaaatgactttggagagaaatataactgtatagcccccaatgactgtgaga 3840
QY 3841 acatcgttcgagagcccaagacagatgatagccaagcccgatgacccaatatacaaa 3900
DB 3841 acatcgttcgagagcccaagacagatgatagccaagcccgatgacccaatatacaaa 3900
QY 3901 gctcgtctctcaagagaggtacacttgcctccacagcagctctccacacctgtcac 3960
DB 3901 gctcgtctctcaagagaggtacacttgcctccacagcagctctccacacctgtcac 3960

DB 3901 gctcgtctctcaagagaggtacacttgcctccacagcagctctccacacctgtcac 3960
QY 3961 tctcgtctcaagagagctatgaacagatgtatgtcattcaactatgtatgag 4020
DB 3961 tctcgtctcaagagagctatgaacagatgtatgtcattcaactatgtatgag 4020
QY 4021 agactggggtccggaagagaccttaacagacatccgctgaactgtgagagagtgtaacctg 4080
DB 4021 agactggggtccggaagagaccttaacagacatccgctgaactgtgagagagtgtaacctg 4080
QY 4081 tcaactgacgaagagcagatgtgtgtgaaggtgaagaggtgtggtggcgagtggtg 4140
DB 4081 tcaactgacgaagagcagatgtgtgtgaaggtgaagaggtgtggtggcgagtggtg 4140
QY 4141 aggggacccagagatcttggtccagagccttgccgaagaaagaaactcaagcagaagctg 4200
DB 4141 aggggacccagagatcttggtccagagccttgccgaagaaagaaactcaagcagaagctg 4200
QY 4201 agctcatggtccctcaagcagcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4260
DB 4201 agctcatggtccctcaagcagcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4260
QY 4261 cggacagcaggtatgtccttggtgactgacagatcaacaggggcatatataaacaagg 4320
DB 4261 cggacagcaggtatgtccttggtgactgacagatcaacaggggcatatataaacaagg 4320
QY 4321 ggttgctaccccaagagaggaagaaataaagaaagaaagaaatctaaagcattag 4380
DB 4321 ggttgctaccccaagagaggaagaaataaagaaagaaagaaatctaaagcattag 4380
QY 4381 aagccttcaatttgcccaaaagctgaactatatacacttccctgagacatcaagcattag 4440
DB 4381 aagccttcaatttgcccaaaagctgaactatatacacttccctgagacatcaagcattag 4440
QY 4441 aagatctcaatctcagagaggaacagagtgctgaacgggttgccaagcagagccagag 4500
DB 4441 aagatctcaatctcagagaggaacagagtgctgaacgggttgccaagcagagccagag 4500
QY 4501 ctgttcaaccttgcctataataagaaagcccaaaagcccaagacccagagacagatga 4560
DB 4501 ctgttcaaccttgcctataataagaaagcccaaaagcccaagacccagagacagatga 4560
QY 4561 ccttagaagagctgcaagagataaagaaatagaccagttctctgaagatcccgagagga 4620
DB 4561 ccttagaagagctgcaagagataaagaaatagaccagttctctgaagatcccgagagga 4620
QY 4621 cctgtatatacctataatggaagaaatccctgcccacaagaaaggtttagaatagtcc 4680
DB 4621 cctgtatatacctataatggaagaaatccctgcccacaagaaaggtttagaatagtcc 4680
QY 4681 aacagatatacctataatggaagaaatccctgcccacaagaaaggtttagaatagtcc 4740
DB 4681 aacagatatacctataatggaagaaatccctgcccacaagaaaggtttagaatagtcc 4740
QY 4741 ccccttatacagttctgaagagcttaccagaggtgtgactcctcagcaggttgtgaagact 4800
DB 4741 ccccttatacagttctgaagagcttaccagaggtgtgactcctcagcaggttgtgaagact 4800
QY 4801 cctgcaagctgttgaatgtatgtatccctcagaatccctcagaagagagactcaaggagaa 4860
DB 4801 cctgcaagctgttgaatgtatgtatccctcagaatccctcagaagagagactcaaggagaa 4860
QY 4861 gccacccaagcgtcactggaaggtgagactcaatgaaggttaagccgcttaatacggaa 4920
DB 4861 gccacccaagcgtcactggaaggtgagactcaatgaaggttaagccgcttaatacggaa 4920
QY 4921 acaataatcatgtgtttttgtagaacacttccagaggtgtgtagaggttatctacta 4980
DB 4921 acaataatcatgtgtttttgtagaacacttccagaggtgtgtagaggttatctacta 4980
QY 4981 agaaagagacttcaacgtgtgtgtgaagaaatactgtgagagaaatttccagattg 5040
DB 4981 agaaagagacttcaacgtgtgtgtgaagaaatactgtgagagaaatttccagattg 5040

Dh	4981	agaaagagccttcaaccgtgtgtgtgtgtaagaaatactgtgaggaatttttccaaagtctg	5040
QY	5041	gaatacctaagrraatataggttcagacaatgtgtccagcttctgtgtgccaggtatgctcag	5100
Dh	5041	gaatacctaagrraatataggttcagacaatgtgtccagcttctgtgtgccaggtatgctcag	5100
QY	5101	gactgtgcgaagatattgtgggattgtattgtggaaactgtcatgtgtcaatacagccccaaagt	5160
Dh	5101	gactgtgcgaagatattgtgggattgtattgtggaaactgtcatgtgtcaatacagccccaaagt	5160
QY	5161	caggacaggttagagagagatgatataagaccattaaagagagacccttaaccaattgtccagag	5220
Dh	5161	caggacaggttagagagagatgatataagaccattaaagagagacccttaaccaattgtccagag	5220
QY	5221	agactgtgcataatgatattgtgatgtgtctctcgtccctttgtgtttttaggggttagaagaa	5280
Dh	5221	agactgtgcataatgatattgtgatgtgtgtctctcgtccctttgtgtttttaggggttagaagaa	5280
QY	5281	ccccgtgcagatttggtgtgtgtgccccctatgaaattgtctctacgggtgagagccccccgtgtg	5340
Dh	5281	ccccgtgcagatttggtgtgtgtgccccctatgaaattgtctctacgggtgagagccccccgtgtg	5340
QY	5341	cccgatgtgcctttgcacatagtctctatgtgtgtgtgtcttcccgagcctttgtctctatggc	5400
Dh	5341	cccgatgtgcctttgcacatagtctctatgtgtgtgtgtcttcccgagcctttgtctctatggc	5400
QY	5401	tcaagagcgtctcgagtgtgggtgagcgagcgtgtgaaagcagctcccgagagccttaatacag	5460
Dh	5401	tcaagagcgtctcgagtgtgggtgagcgagcgtgtgaaagcagctcccgagagccttaatacag	5460
QY	5461	gaaggaagactgtgcagatgtccacatactgcttcccaagtgtgagatctcagtctatgataagccgc	5520
Dh	5461	gaaggaagactgtgcagatgtccacatactgcttcccaagtgtgagatctcagtctatgataagccgc	5520
QY	5521	accgtgtcaggaataacccctgcagactcgtgtgaaagagacctaactatctcgttaactttgacacac	5580
Dh	5521	accgtgtcaggaataacccctgcagactcgtgtgaaagagacctaactatctcgttaactttgacacac	5580
QY	5581	caacgtgtctgtgaaagttcgaagaaatccccaactgtatccatgtgacatcccaagttaaagccgg	5640
Dh	5581	caacgtgtctgtgaaagttcgaagaaatccccaactgtatccatgtgacatcccaagttaaagccgg	5640
QY	5641	cgccaactcccgactcgtgggtgtgagagccgaaagacctaagatccccaactgaagcttgcgc	5700
Dh	5641	cgccaactcccgactcgtgggtgtgagagccgaaagacctaagatccccaactgaagcttgcgc	5700
QY	5701	tccatcgtcgtgttctcactaacttaacaaataaccccccaagccaggtatgaacgccttata	5760
Dh	5701	tccatcgtcgtgttctcactaacttaacaaataaccccccaagccaggtatgaacgccttata	5760
QY	5761	gacagctcgaacccccatagacacttatacccttaactcctgtgtatattgacccctgtatag	5820
Dh	5761	gacagctcgaacccccatagacacttatacccttaactcctgtgtatattgacccctgtatag	5820
QY	5821	ggtgtcactgttaataagcaccctcgagatgtgtgtctcctatgagagcaccctgtgtccctgtacgt	5880
Dh	5821	ggtgtcactgttaataagcaccctcgagatgtgtgtgtctcctatgagagcaccctgtgtccctgtacgt	5880
QY	5881	catcttcgtcctccgattgtataaccgcgtgtttaaagcaacaccccccaactagatcgt	5940
Dh	5881	catcttcgtcctccgattgtataaccgcgtgtttaaagcaacaccccccaactagatcgt	5940
QY	5941	agttatgtgtttctattgtgtcccaagcacaagaaagaaataactgtgtgggttctcgg	6000
Dh	5941	agttatgtgtttctattgtgtcccaagcacaagaaagaaataactgtgtgggttctcgg	6000
QY	6001	gaatcctctgttagagagatgtgagcgtgtgtcactctccaagatgtgagactgtgaaatgtgcg	6060
Dh	6001	gaatcctctgttagagagatgtgagcgtgtgtcactctccaagatgtgagactgtgaaatgtgcg	6060
QY	6061	atctcttcacagagccgggtaaataattctccttgtcaattccgcccgggcaagtatacaa	6120
Dh	6061	atctcttcacagagccgggtaaataattctccttgtcaattccgcccgggcaagtatacaa	6120

QY	6121	atgataaactatataaagataagagctgtctccccaatcagaacttagttttcttaaaagata	6180
Db	6121	atgataaactatataaagataagagctgtctccccaatcagaacttagttttcttaaaagata	6180
QY	6181	agttccatcgaagaaacagaaatatttcaaaagctgaataatgttatgactgtggg	6240
Db	6181	agttccatcgaagaaacagaaatatttcaaaagctgaataatgttatgactgtggg	6240
QY	6241	aattgtttttattatataatgctggggagacaggggtccactttaacattcgcttaggata	6300
Db	6241	aattgtttttattatataatgctggggagacaggggtccactttaacattcgcttaggata	6300
QY	6301	gagacggggacagaacccccctgtgacaaatgggccccgaataaagctactgtgtgaacagggg	6360
Db	6301	gagacggggacagaacccccctgtgacaaatgggccccgaataaagctactgtgtgaacagggg	6360
QY	6361	cccccgccccctgtgagccacacgcatactgtgcggtgtccccaattaaacctgtgtgcgct	6420
Db	6361	cccccgccccctgtgagccacacgcatactgtgcggtgtccccaattaaacctgtgtgcgct	6420
QY	6421	gacataacacagccgcttagacaaacgataccactgtgatttgaatttcctaccaaagcgttga	6480
Db	6421	gacataacacagccgcttagacaaacgataccactgtgatttgaatttcctaccaaagcgttga	6480
QY	6481	aactccccaggtgttcctgttaagacaaagagactcttcagttccatccagggagct	6540
Db	6481	aactccccaggtgttcctgttaagacaaagagactcttcagttccatccagggagct	6540
QY	6541	ttcccaagccatacaactccacacgacctgtatgccaattctctgttgcctttttatccc	6600
Db	6541	ttcccaagccatacaactccacacgacctgtatgccaattctctgttgcctttttatccc	6600
QY	6601	tcagggccctcttattatagagggatgtgtgttaagaagaagaatccaatgtatgaccaagag	6660
Db	6601	tcagggccctcttattatagagggatgtgtgttaagaagaagaatccaatgtatgaccaagag	6660
QY	6661	catagaatacatgtatagatgggggtcccgaaataagcgttaacctcagatgaattccggg	6720
Db	6661	catagaatacatgtatagatgggggtcccgaaataagcgttaacctcagatgaattccggg	6720
QY	6721	aaggggacatgcatagagaaagctcccccatacccaacacaccttgctatagtaactgtg	6780
Db	6721	aaggggacatgcatagagaaagctcccccatacccaacacaccttgctatagtaactgtg	6780
QY	6781	gtttatgagcagggctcagaaatccagtatttgatgacctggttataacaggtgtgtggca	6840
Db	6781	gtttatgagcagggctcagaaatccagtatttgatgacctggttataacaggtgtgtggca	6840
QY	6841	tgaataactgggtttiaacccccctgtgttccacctcagttccatccaacccaataaagtttc	6900
Db	6841	tgaataactgggtttiaacccccctgtgttccacctcagttccatccaacccaataaagtttc	6900
QY	6901	tgtgtcatggtcccaatcgtcccccagatgtgtactataccatctctgaggaagtgctctgtat	6960
Db	6901	tgtgtcatggtcccaatcgtcccccagatgtgtactataccatctctgaggaagtgctctgtat	6960
QY	6961	gaatatgactatcggtatataccgacccaataaagagaaacccgfatatcccttacccttagctgta	7020
Db	6961	gaatatgactatcggtatataccgacccaataaagagaaacccgfatatcccttacccttagctgta	7020
QY	7021	atgctcgatattagagacggcgttgcgtgtagtagaacaagagacagctgtccctgtatcaagga	7080
Db	7021	atgctcgatattagagacggcgttgcgtgtagtagaacaagagacagctgtccctgtatcaagga	7080
QY	7081	ccacagcagctatagaaagagacttgtgtgctgtacatgtggccatbaacgaagaactctcga	7140
Db	7081	ccacagcagctatagaaagagacttgtgtgctgtacatgtggccatbaacgaagaagaactctcga	7140
QY	7141	gacctaaaggagctgtttagaaccttagaacctgaagagcttcttctgtctgaagtgtgt	7200
Db	7141	gacctaaaggagctgtttagaaccttagaacctgaagagcttcttctgtctgaagtgtgt	7200

```

QY 7201 ctacagacccgagggaggtatagatctgtcttctaaagaagtggtatgtcagcc 7260
DB 7201 ctacagacccgagggaggtatagatctgtcttctaaagaagtggtatgtcagcc 7260
QY 7261 ttaagaagaatgtctctctatgtatgactcagggagccatcagaactcctagac 7320
DB 7261 ttaagaagaatgtctctctatgtatgactcagggagccatcagaactcctagac 7320
QY 7321 aagcttagaaaaaagttagagagcgctcgaaggaaagagagagcgctcaggggttt 7380
DB 7321 aagcttagaaaaaagttagagagcgctcgaaggaaagagagagcgctcaggggttt 7380
QY 7381 gaagagatgtctcaacagctctccttgatgacacccctgtctctgtcctgaagggcc 7440
DB 7381 gaagagatgtctcaacagctctccttgatgacacccctgtctctgtcctgaagggcc 7440
QY 7441 ctatagctctctctctgttacttaacgttgagccttgcttctaaatagattgttgc 7500
DB 7441 ctatagctctctctctgttacttaacgttgagccttgcttctaaatagattgttgc 7500
QY 7501 ttgttagagagacgagtgatgacagctcagctatggttacttaggcaacagtaacagc 7560
DB 7501 ttgttagagagacgagtgatgacagctcagctatggttacttaggcaacagtaacagc 7560
QY 7561 ctctgagccagagagaaactgacctctagcctccagctcctcaagattagaactatla 7620
DB 7561 ctctgagccagagagaaactgacctctagcctccagctcctcaagattagaactatla 7620
QY 7621 caagacaagaagtgagggaatgaaagtgaatgaatgaatgaatgaatgaatgaatgaat 7680
DB 7621 caagacaagaagtgagggaatgaaagtgaatgaatgaatgaatgaatgaatgaatgaat 7680
QY 7681 aagtaataaagaactcctaaatgccccgaatgaatgaatgaatgaatgaatgaatgaat 7740
DB 7681 aagtaataaagaactcctaaatgccccgaatgaatgaatgaatgaatgaatgaatgaat 7740
QY 7741 ggtagaaggtcacactctcctatgttccagggcctgctacccctgagctgaatgaat 7800
DB 7741 ggtagaaggtcacactctcctatgttccagggcctgctacccctgagctgaatgaat 7800
QY 7801 aggaatgagttgactaactcgtactcgtactcgtactcgtactcgtactcgtactcgt 7860
DB 7801 aggaatgagttgactaactcgtactcgtactcgtactcgtactcgtactcgtactcgt 7860
QY 7861 aattgattacacatgagacccctagtgactatcctaaactgaatcgtactcgtactcgt 7920
DB 7861 aattgattacacatgagacccctagtgactatcctaaactgaatcgtactcgtactcgt 7920
QY 7921 aggaagccacagagatgagacccctcagagctatcttaaatgattggtccacagagcc 7980
DB 7921 aggaagccacagagatgagacccctcagagctatcttaaatgattggtccacagagcc 7980
QY 7981 gggctcctgatatcttaaatgattggtccacagagcgggctcgtactcgtactcgtact 8040
DB 7981 gggctcctgatatcttaaatgattggtccacagagcgggctcgtactcgtactcgtact 8040
QY 8041 gattggtttgtgacgacagagcttgttgaaccccaataaagttgcccagatccgca 8100
DB 8041 gattggtttgtgacgacagagcttgttgaaccccaataaagttgcccagatccgca 8100
QY 8101 ctccgggacgagctcctcctcagcctgtggtgtagactgtgggcccacagcgttgg 8160
DB 8101 ctccgggacgagctcctcctcagcctgtggtgtagactgtgggcccacagcgttgg 8160
QY 8161 aataaaatcctctgtctgttgcatacaaaaaaataaaaaaataaaaaa 8209
DB 8161 aataaaatcctctgtctgttgcatacaaaaaaataaaaaaataaaaaa 8209

```

RESULT 2
AAV09699
ID AAV09699 standard: DNA: 8196 BP.
HX

```

AC AAV09699;
XX
DT 19-MAY-1998 (first entry)
XX
DE Porcine retrovirus DNA encoding, GAG, POL and ENV.
XX
KW Porcine retrovirus: PoRV; POL protein; ENV protein; GAG protein;
XX vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
OS Porcine retrovirus.
XX
FH
FT Key Location/Qualifiers
FT CDS 576..2126
FT /tag= a
FT /product= GAG protein
FT /note= "viroin core polypeptide"
FT 2143..5733
FT /tag= b
FT /product= POL protein
FT /note= "polymerase peptide sequence as given in
FT specification"
FT 5606..7576
FT CDS
FT /tag= c
FT /product= ENV protein
FT /note= "envelope protein"
XX
PN MO9740167-A1.
XX
PD 30-OCT-1997.
XX
PF 18-APR-1997; 97MO-GB01087.
XX
PR 10-FEB-1997; 97GB-0002668.
XX 19-APR-1996; 96GB-0008164.
XX
PA (IMUT-) IMUTRAN LTD.
XX (QONE-) Q-ONE BIOTECH LTD.
XX
PI Galbraith DN, Haworth C, Lees GM, Smith KT;
XX WPI: 1997-535851/49.
XX
DR
XX
PT Polynucleotide encoding porcine retrovirus expression product -
XX useful to develop products for use in vaccines, diagnosis and
XX xeno-transplantation
XX
PS Claim 4: Fig 2: 69pp: English.
XX
CC This DNA sequence encodes the porcine retrovirus (PoRV) viroin core
CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
CC proteins can be used to develop viral vaccines, antisense nucleic acids,
CC ribozymes and other antiviral agents. They can also be used in
CC xeno-transplantation technology and as diagnostic tools.
XX
SQ Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other.

```

Query Match 95.0%; Score 7802; DB 18; Length 8196;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 8192; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```

QY 13 tgtggcccccggcgcttggataaaatcctctgtgtgtgtgatacaagcgttct 72
DB 1 tgtggcccccggcgcttggataaaatcctctgtgtgtgtgatacaagcgttct 72
QY 73 cgtgagtgatttgggtgtgtcctcttcgagccggagcagaggggattgtcttact 132
DB 61 cgtgagtgatttgggtgtgtcctcttcgagccggagcagaggggattgtcttact 132
QY 133 ggccttcatttgggtgtgtccttcgagccgggaaatcctcgcacaccccttaacacggagac 192
DB 121 ggccttcatttgggtgtgtccttcgagccgggaaatcctcgcacaccccttaacacggagac 180

```

QY	193	gacttggaggtgaagggatccctcttggaaagtgtgtgttcggtcggtctctgtt	252
Db	181	gacttggaggtgaagggatccctcttggaaagtgtgtgttcggtcggtctctgtt	240
QY	253	ctgagtgatcttctgtgtatcggtcttctgttgcagctgtctctcagaccgttaa	312
Db	241	ctgagtgatcttctgtgtatcggtcttctgttgcagctgtctctcagaccgttaa	300
QY	313	ggaactggagactctgtatcagcaagactgtcattgagagatcaagactgccacctgggg	372
Db	301	ggactggagagactgtatcagcaagactgtcattgagagatcaagagctgtccacctgggg	360
QY	373	acgcccggagagtggtggagagccagggagacgctgtgtgtctctactgtctgattcagag	432
Db	361	acgcccggagagtggtggagagccagggagacgctgtgtgtctctactgtctgattcagag	420
QY	433	accgagttctgtgtgtgaagcgaaagcttccctccggtcggtccgactcttggct	492
Db	421	accgagttctgtgtgtgaagcgaaagcttccctccggtcggtccgactcttggct	480
QY	493	gcttgttgaaagacgcggagcgggttcgctgtgtcttgatctgttggttctgtctcgtgtg	552
Db	481	gcttgttgaaagacgcggagcgggttcgctgtgtgtcttgatctgttggttctgtctcgtgtg	540
QY	553	tcttcttctgtgtgcgtctctgtctatacatttaataatgtagaacaagttgactaccccc	612
Db	541	tcttcttctgtgtgcgtctctgtctatacagtttaataatgtagaacaagttgactaccccc	600
QY	613	ttagttgactctcgaaccattgtgactggaagttaagttccagggctataattgttcagtctc	672
Db	601	ttagttgactctcgaaccattgtgactggaagttaagttccagggctataattgttcagtctc	660
QY	673	aggttaagagggagacctctgacagacttctgtgcctctgaatgtgcacaactcgaatttg	732
Db	661	aggttaagagggagacctctgacagacttctgtgcctctgaatgtgcacaactcgaatttg	720
QY	733	gatgtgcatcagagggagacctttaatcttcgaaattatccctgtctttaaagcaatcatt	792
Db	721	gatgtgcatcagagggagacctttaatcttcgaaattatccctgtctttaaagcaatcatt	780
QY	793	ttcagacttgagcccggtctctcatccttgatataagagacccatatccttaacgtgtgcaagatt	852
Db	781	ttcagacttgagcccggtctctcatccttgatataagagacccatatccttaacgtgtgcaagatt	840
QY	853	tgtcagaagatctctccgccaatgtggttaaaccaatggtctaaataaacaaagagccaggtc	912
Db	841	tgtcagaagatctctccgccaatgtggttttaaaccaatggtctaaataaacaaagagccaggtc	900
QY	913	cccggaatccctgtccttcttgagagagaaaaacaacactgtgcggaaaaagtctgagccctctt	972
Db	901	cccggaatccctgtccttcttgagagagaaaaacaacactgtgcggaaaaagtctgagccctctt	960
QY	973	cctctgatctaaccccagatctcgagagagccgagactgtgcggaaaaccccaacctgtcc	1032
Db	961	cctctgatctaaccccagatctcgagagagccgagactgtgcggaaaaccccaacctgtctcc	1020
QY	1033	ccacccccttatccagcaagagtgctctgttgagggagacctctgcccctctgtgagctccgg	1092
Db	1021	ccacccccttatccagcaagagtgctctgttgagggagacctctgcccctctgtgagctccgg	1080
QY	1093	tgtgtgagagagactctgtgcgggactctgagagccggagagagcgcgcaaccccgagagcgag	1152
Db	1081	tgtgtgagagagactctgtgcgggactctgagagccggagagagcgcgcaaccccgagagcgag	1140
QY	1153	acgagatcggtatatcagctgtgcgcacattatggccctcccatctgcagaggggccaattgc	1212
Db	1141	acgagatcggtatatcagctgtgcgcacattatggccctcccatctgcagaggggccaattgc	1200
QY	1213	agccctccagatattggcccttctctctgcagatctctataattgaaaaactaacctc	1272
Db	1201	agccctccagatattggcccttctctctgcagatctctataattgaaaaactaacctc	1260
QY	1273	cccccttctcgagagatccccaacgctcaagggtgttgtagtgccttatgttctctc	1332

Db	1261	cccccttctcgagagatcccccaacgctccacggtgttctggtgagatcccttaagttctctc	1320
Oy	1333	accagcctactctgggaatgattgtcaaaagctgtcgcagacaccttccacaacccgaggaagc	1392
Db	1321	accagcctactctgggaatgattgtcaaaagctgtcgcagacaccttccacaacccgaggaagc	1380
Oy	1393	gagagagagatctcgtcttagagcgtctgaataaaatgltctctgggcccgaacggcgacccagc	1452
Db	1381	gagagagagatctcgtcttagagcgtctgaataaaatgltctctgggcccgaacggcgacccagc	1440
Oy	1453	agttgcaaaaatgagattgacatgggatttcccttgactcgcgcccggtctggagatacaaca	1512
Db	1441	agttgcaaaaatgagattgacatgggatttcccttgactcgcgcccggtctggagatacaaca	1500
Oy	1513	cggctgaagatgtagagagagctgtgaaaatctatcgccaggtctcgtgtgcggtgtctccggg	1572
Db	1501	cggctgaagatgtagagagagctgtgaaaatctatcgccaggtctcgtgtgcggtgtctccggg	1560
Oy	1573	ggcgctcaagaacggccccaataattggctcaaggtaagaaggtatccagagacccgaacg	1632
Db	1561	ggcgctcaagaacggccccaataattggctcaaggtaagaaggtatccagagacccgaacg	1620
Oy	1633	aacctccctcgatcttctcttgagaggtctcatgaaagcctcagcggtctaacccctttg	1692
Db	1621	aacctccctcgatcttctcttgagaggtctcatgaaagcctcagcggtctaacccctttg	1680
Oy	1693	atccctacccagaagccccaagaagctctcaggttgcccttgcttcatcttgagatctcgctc	1752
Db	1681	atccctacccagaagccccaagaagctctcaggttgcccttgcttcatcttgagatctcgctc	1740
Oy	1753	tgggtatcaagaagaagaactlcaagagcttgaagaggtttacagaagcctgaatgaatcgtgac	1812
Db	1741	tgggtatcaagaagaagaactlcaagagcttgaagaggtttacagaagcctgaatgaatcgtgac	1800
Oy	1813	tagtgaagagagcgagagagaagctgtatatacagaagggagacagaagaagggagagagacga	1872
Db	1801	tagtgaagagagcgagagagaagctgtatatacagaagggagacagaagaagggagagagacga	1860
Oy	1873	gaaaaagaaaagagagagagaagaagaagggagagaaagacgtgataagcggcagaagagaatt	1932
Db	1861	gaaaaagaaaagagagagagaagaagaagggagagaaagacgtgataagcggcagaagagaatt	1920
Oy	1933	tgactaaagatcttggccgacagtgtgtgaagggagagcgacagcgagggagagagaagatct	1992
Db	1921	tgactaaagatcttggccgacagtgtgtgaagggagagcgacagcgagggagagagaagatct	1980
Oy	1993	ttagagaaaatttaggttcagggccctcagacagttacagaaagcctcggggcaatagaagccccatcg	2052
Db	1981	ttagagaaaatttaggttcagggccctcagacagttacagaaagcctcggggcaatagaagccccatcg	2040
Oy	2053	acaaagagaccagtgctgctatgttaagaagaagaagacactgycgcgaaggaatctgcccaga	2112
Db	2041	acaaagagaccagtgctgctatgttaagaagaagaagacactgycgcgaaggaatctgcccaga	2100
Oy	2113	aagggaaaacaagagacccgaag-gtccacagctcagaagaagattaaagatttaaggggagac	2172
Db	2101	aagggaaaacaagagacccgaag-gtccacagctcagaagaagattaaagatttaaggggagac	2158
Oy	2173	gttcgcgacccctcccccagcccgacaggttaacttgaaggttgaggggcaaccagttgaat	2232
Db	2159	gttcgcgacccctcccccagcccgacaggttaacttgaaggttgaggggcaaccagttgaat	2218
Oy	2233	tccctgtgtgataccggagcggtgagcttaaggttcgtctacaacccattagaanaaatlaaag	2292
Db	2219	tccctgtgtgataccggagcggtgagcttaaggttcgtctacaacccattagaanaaatlaaag	2278
Oy	2293	aaaaaaaatcctgtgtgatacgggtgtccacagggccaacggaatccatgagataccgacccga	2352
Db	2279	aaaaaaaatcctgtgtgatacgggtgtccacagggccaacggaatccatgagataccgacccga	2338
Oy	2353	gaacggttagacttggaggtgggacgggttaaccacgttcttggtatccctgaatctcc	2412

Db 2339 gaacggttgacttggtgagtgagacgggtacccaactcgtttctgtgcatcccttgatgcc 2398
QY 2413 cagtcacctctagtagagactactgacaaagatgagagcctcaatttttggac 2472
Db 2399 cagtaacctctctagtagagactactgacaaagatgagagcctcaatttttggac 2458
QY 2473 aaggaagaccgaaggtgtgtgtgataaacaacccaactactgtgttgaacctcaatag 2532
Db 2459 aaggaagaccgaaggtgtgtgtgataaacaacccaactactgtgttgaacctcaatag 2518
QY 2533 atgataatatacgaactatctctcccaagtaagacctggttcaagtatacagtcctgt 2592
Db 2519 atgataatatacgaactatctctcccaagtaagacctggttcaagtatacagtcctgt 2578
QY 2593 tggagcaggtttccccaagccttggtgagaaacccgcaaggatgggttgcgaagcaagttc 2652
Db 2579 tggagcaggtttccccaagccttggtgagaaacccgcaaggatgggttgcgaagcaagttc 2638
QY 2653 ccccaacagttatcaactgaagcaggtgtacacacagttacagttacagtaacagttac 2712
Db 2639 ccccaacagttatcaactgaagcaggtgtacacacagttacagttacagtaacagttac 2698
QY 2713 tgggtgagagagctcgagagagaaatttggcgcagtttcaagataatcccaacaggca 2772
Db 2699 tgggtgagagagctcgagagagaaatttggcgcagtttcaagataatcccaacaggca 2758
QY 2773 tctctgtctcgtccatcccttggaatccctctgctacacaggttgaagagccttgga 2832
Db 2759 tctctgtctcgtccatcccttggaatccctctgctacacaggttgaagagccttgga 2818
QY 2833 ccaatgattatcgaccacagtaacagacttgagagagcttgaataaagggtgcagagacac 2892
Db 2819 ccaatgattatcgaccacagtaacagacttgagagagcttgaataaagggtgcagagacac 2878
QY 2893 acccaacggtccccaacacctataacctctgagcgcctccgcgtgaacaggaaactgt 2952
Db 2879 acccaacggtccccaacacctataacctctgagcgcctccgcgtgaacaggaaactgt 2938
QY 2953 acacagtttgacttaaaagaatgctctctctgctgagagattacacccaactagccaac 2998
Db 2939 acacagtttgacttaaaagaatgctctctctgctgagagattacacccaactagccaac 3012
QY 3013 caattttgctctcgaattgagagatccaggttcaggaagacccgggagcagttcaacctga 3072
Db 2999 caattttgctctcgaattgagagatccaggttcaggaagacccgggagcagttcaacctga 3058
QY 3073 ccgagacgcccgaaggttcaagaactcccgacacactcttgaagagaccccaacaggg 3118
Db 3059 ccgagacgcccgaaggttcaagaactcccgacacactcttgaagagaccccaacaggg 3132
QY 3133 accgtgcgaactcagagatcccaacacctcaggttcaggttcacctcctcagttcagttgac 3192
Db 3119 accgtgcgaactcagagatcccaacacctcaggttcaggttcacctcctcagttcagttgac 3178
QY 3193 tgccttcggtgagagacccaacaaacagagactgttgaaggttcaggaagcactactgcgtg 3252
Db 3179 tgccttcggtgagagacccaacaaacagagactgttgaaggttcaggaagcactactgcgtg 3238
QY 3253 aattgtctgacactgacagactcgtctgaagaagccagatttgcagagagagag 3312
Db 3239 aattgtctgacactgacagactcgtctgaagaagccagatttgcagagagagag 3298
QY 3313 taacatacttgggtgacagtttgcgggagcgggagtggttcgaagagcagcagaaga 3372
Db 3299 taacatacttgggtgacagtttgcgggagcgggagtggttcgaagagcagcagaaga 3358
QY 3373 aaactgtgtccagatcccgcccaacacagcagcagaagtaagagagatttttggga 3432
Db 3359 aaactgtgtccagatcccgcccaacacagcagcagaagtaagagagatttttggga 3418
QY 3433 cagctggaatttgcagagatggttccggggttgcagaccttaagcccaactac 3492
Db 3419 cagctggaatttgcagagatggttccggggttgcagaccttaagcccaactac 3478

QY 3493 cgtacaacaaagaaagggggttctccgtggtctcgtgagcaccagaagcatttgaag 3552
Db 3479 cgtacaacaaagaaagggggttctccgtggtctcgtgagcaccagaagcatttgaag 3538
QY 3553 ctatcaaaaagccttgctgtagcagactgtctctggtggtccctctgagcactaaacct 3612
Db 3539 ctatcaaaaagccttgctgtagcagactgtctctggtggtccctctgagcactaaacct 3598
QY 3613 ttaccttttgtgtgagagcgttaaggagtagcccgagaggttttaaccacaacctag 3672
Db 3599 ttaccttttgtgtgagagcgttaaggagtagcccgagaggttttaaccacaacctag 3658
QY 3673 gaacatgagagagacccgtgtgtctacactgttcaaaagagtttaccctgtagcagttgt 3732
Db 3659 gaacatgagagagacccgtgtgtctacactgttcaaaagagtttaccctgtagcagttgt 3718
QY 3733 ggtccgtatgtctgaaagcgtatccgagctgtgtgacatactgttcaagagcgtgacaat 3792
Db 3719 ggtccgtatgtctgaaagcgtatccgagctgtgtgacatactgttcaagagcgtgacaat 3778
QY 3793 tgaatttggagacagataataactgttaataagcccccatgtcatgtgagaaatcgttcg 3852
Db 3779 tgaatttggagacagataataactgttaataagcccccatgtcatgtgagaaatcgttcg 3838
QY 3853 agcccccaagaccgatgtgatacacaacgcccagatgacacatatacaagcctgttctca 3912
Db 3839 agcccccaagaccgatgtgatacacaacgcccagatgacacatatacaagcctgttctca 3898
QY 3913 cagagaggttcaacttctgctccacacagcgtctcaacacctgtcacactctctgtgaag 3972
Db 3899 cagagaggttcaacttctgctccacacagcgtctcaacacctgtgtcacactctctgtgaag 3958
QY 3973 agactgtgaaacagtgatcatatgtatgtccatcaactatgtatgtgagagagcgtgtcc 4032
Db 3959 agactgtgaaacagtgatcatatgtatgtccatcaactatgtatgtgagagagcgtgtcc 4018
QY 4033 gcaagacacttaacagataacacgtctgactgtgagaaagtgttaacctgttcaactgacga 4092
Db 4019 gcaagacacttaacagataacacgtctgactgtgagaaagtgttaacctgttcaactgacga 4078
QY 4093 gcaactgtgtgtgtgaaagttaagagatgtgtgtgagcagatgtgtgtgagcagaccca 4152
Db 4079 gcaactgtgtgtgtgaaagttaagagatgtgtgtgagcagatgtgtgtgagcagaccca 4138
QY 4153 cgaactgtgtgtgtgaaagttaagagatgtgtgtgagcagatgtgtgtgagcagaccca 4212
Db 4139 cgaactgtgtgtgtgaaagttaagagatgtgtgtgagcagatgtgtgtgagcagaccca 4198
QY 4213 tcaacgaagcttgcgtgtgagcagaggaatccataaacaatttatacgaacagcaggt 4272
Db 4199 tcaacgaagcttgcgtgtgagcagaggaatccataaacaatttatacgaacagcaggt 4258
QY 4273 atgaccttgcgtgtgagcagcgttacaacgggccaactataaacaaggggtgttacct 4332
Db 4259 atgaccttgcgtgtgagcagcgttacaacgggccaactataaacaaggggtgttacct 4318
QY 4333 cagcagggaggggaataaagaacaaaggagaaattcraagcctttagaagccttacct 4392
Db 4319 cagcagggaggggaataaagaacaaaggagaaattcraagcctttagaagccttacct 4378
QY 4393 tgcataaaaggttaacttatatacgttcccttggaatcagaaagcccaagaatccat 4452
Db 4379 tgcataaaaggttaacttatatacgttcccttggaatcagaaagcccaagaatccat 4438
QY 4453 cttaggggaacaaatgtgtcagcgggttgcagagcagcagccagcgtgttaacctc 4512
Db 4439 cttaggggaacaaatgtgtcagcgggttgcagagcagcagccagcgtgttaacctc 4498
QY 4513 tgcctataataagaacgcccagaagcccagaaacagagcagagtagcaccttagaagact 4572
Db 4499 tgcctataataagaacgcccagaagcccagaaacagagcagagtagcaccttagaagact 4558

us-09-171-553b-3.feb22011.rng

Score 2967; DB 22; Length 15537
36.18; Answer Match

Query Match	36.1%	Score 2507	DB 22	Conserved	2
Best Local Similarity	99.8%	Pred. No. 0			
Matches 3507; Conservative	0	Mismatches	6	Indels	2
				Gaps	2

[illegible]

Db	971	tcgagccctc-tcctcgtatctaccccgagatcgaagagccgaccttgcgcggaaacc	1029
QY	1021	caacctgtccccaaccccttaccagacaggtgtctgagggga-ccctgtgcccc	1079
Db	1030	caacctgtccccaaccccttaccagacaggtgtctgaggggaacccctgtgcccc	1089
QY	1080	cctgtagctccggtgtgagggagacctgtcgcgcgggaactccggagccggagagccacac	1139
Db	1090	cctgtagctccggtgtgagggagacctgtcgcgcgggaactccggagccggagagagccacac	1149
QY	1140	ccggagcggagacagagacgagatccgatatccgcgtgcgcacatgtgccccctccatgccca	1199
Db	1150	ccggagcggagacagagagatccgatatccgcgtgcgcacatgtgccccctccatgccca	1209
QY	1200	gggggccaattgcagccccctccagatgtgcccccttctctgcgagatctctataattg	1259
Db	1210	gggggccaattgcagccccctccagatgtgcccccttctctgcgagatctctataattg	1269
QY	1260	aaaactaacatcccccttctcgtgagagatccccaagccctcaaggggtgtggtgagctcc	1319
Db	1270	aaaactaacatcccccttctcgtgagagatccccaagccctcaaggggtgtggtgagctcc	1329
QY	1320	cttatgtcttccacagactactcgtggagatgattgtccaacagctgtctgcagacactcttc	1379
Db	1330	cttatgtcttccacagactactcgtggagatgattgtccaacagctgtctgcagacactcttc	1389
QY	1380	acaaacggagagcggagagagaattctgttagaggtgtagaaaaaatgattctccggggccgac	1439
Db	1390	acaaacggagagcggagagagaattctgttagaggtgtagaaaaaatgattctccggggccgac	1449
QY	1440	gggggacccaacggagattgtgcaaaatgagatctgacatgtggaattctccctgtgactccgggt	1499
Db	1450	gggggacccaacggagattgtgcaaaatgagatctgacatgtggaattctccctgtgactccgggt	1509
QY	1500	tgggactacacaacacggtctgaaggtlaagtgagagctgtgaaaatcattatccgcaggtctgtg	1559
Db	1510	tgggactacacaacacggtctgaaggtlaagtgagagctgtgaaaatcattatccgcaggtctgtg	1569
QY	1560	gcgggtctccgggctgcgctccaaagacggccacacaaattgtgctaaggtaaagaggtgattg	1619
Db	1570	gcgggtctccgggctgcgctccaaagacggccacacaaattgtgctaaggtaaagaggtgattg	1629
QY	1620	caggagaccgaagacaaactccctccgtatctcttgagaggtctcaatggaacccctcaagcgg	1679
Db	1630	caggagaccgaagacaaactccctccgtatctcttgagaggtctcaatggaacccctcaagcgg	1689
QY	1680	ttcaaccccttgtatctcactactccagagggcccaagaacccctcaagctgcgctgtccattt	1739
Db	1690	ttcaaccccttgtatctcactactcccaagggcccaagaacccctcaagctgcgctgtccattt	1749
QY	1740	ggggcagctcggtcgtgatatcagagagaagaactccaagagactgtgaaggtttacaaggaggtc	1799
Db	1750	ggggcagctcggtcgtgatatcagagagaagaactccaagagactgtgaaggtttacaaggaggtc	1809
QY	1800	gaattacgttatctcagtgagagagggcagagaggtgtatccaaagggagagacacgaagag	1859
Db	1810	gaattacgttatctcagtgagagagggcagagaggtgtgtttacagaagggagagacgaagag	1869
QY	1860	gagagaggaacagagaaaagaaagagagagagagagagagagagagagagagagagagag	1919
Db	1870	gagagaggaacagagaaaagaaagagagagagagagagagagagagagagagagagagag	1929
QY	1920	caagagagaagaattgtactaagaatctctggccgcgcaggtgttgaagggaaagagcagcagggag	1979
Db	1930	caagagagaagaattgtactaagaatctctggccgcgcaggtgttgaagggaaagagcagcagggag	1989
QY	1980	agaaagagagagatttttagaagaatttagttcaaggccctagaacatcagtggaacctgtggcaat	2039
Db	1990	agaaagagagagatttttagaagaatttagttcaaggccctagaacatcagtggaacctgtggcaat	2049
QY	2040	agagccccaactcgacaagagccaggtgtggtattgttaagaaaaaagaacatcgggccaag	2099

```

Db 2050 aggagcccaactcgacaaagaccagtgctgattgttaagaaaaagacactgagcag 2109
QY 2100 aactgcccagaagaagggaacaagaagccgaaggtccttaacttaagaagaataaagat 2159
Db 2110 aactgcccagaagaagggaacaagaagccgaaggtccttaacttaagaagaataaagat 2169
QY 2160 taggggaagcgggttcggaccccccccgagcccgaggttaactttggaaggtggaagg 2219
Db 2170 taggggaagcgggttcggaccccccccgagcccgaggttaactttggaaggtggaagg 2229
QY 2220 caacacgttgatctcgtgttgatccggagcggaacatcagtgctgtacaacatla 2279
Db 2230 caacacgttgatctcgtgttgatccggagcggaacatcagtgctgtacaacatla 2289
QY 2280 ggaataactaaagaaaaaaatccctggtgtaagtggtgccaacgggcaacggcagatcca 2339
Db 2290 ggaataactaaagaaaaaaatccctggtgtaagtggtgccaacgggcaacggcagatcca 2349
QY 2340 tggactacccgaagaacccgttgacttggaagtggtggaacgggtaacccactcgttcgtc 2399
Db 2350 tggactacccgaagaacccgttgacttggaagtggtggaacgggtaacccactcgttcgtc 2409
QY 2400 atccctgagtgcccaagtaacccctctaggtagaagacttaactgacccaagaatggagctcaa 2459
Db 2410 atccctgagtgcccaagtaacccctctaggtagaagacttaactgacccaagaatggagctcaa 2469
QY 2460 attcttctgaaacaagaagaccagaatgtctgtgaatacaaacaccactactgtgtg 2519
Db 2470 attcttctgaaacaagaagaccagaatgtctgtgaatacaaacaccactactgtgtg 2529
QY 2520 accctccaattagatgataatgcataatctctcccaagtaagcctgtgtaagat 2579
Db 2530 accctccaattagatgataatgcataatctctcccaagtaagcctgtgtaagat 2589
QY 2580 atacagtcctggttggaacagttccccaagcctgggacagaacccgacaggaatgggtg 2639
Db 2590 atacagtcctggttggaacagttccccaagcctgggacagaacccgacaggaatgggtg 2649
QY 2640 gcaagagcaagtcctcccaacaggtatcaactgaagcgaagtgcttaacacatcagtc 2699
Db 2650 gcaagagcaagtcctcccaacaggtatcaactgaagcgaagtgcttaacacatcagtc 2709
QY 2700 agacagtaacccctggtgagtagagaggtcgaagaagaattggccgcatgttcaagaatla 2759
Db 2710 agacagtaacccctggtgagtagagaggtcgaagaagaattggccgcatgttcaagaatla 2769
QY 2760 atccaacagggcacttctcgtgccaatccctctggaatcactccctcgtcaccggtt 2819
Db 2770 atccaacagggcacttctcgtgccaatccctctggaatcactccctcgtcaccggtt 2829
QY 2820 aggaagccttggaacacatataatgcacagtaacagacttgagaagagtgcaataaagg 2879
Db 2830 aggaagccttggaacacatataatgcacagtaacagacttgagaagagtgcaataaagg 2889
QY 2880 gttcagagataataacccaacgggtcccgaaaccttataaacccttgagccctccgcct 2939
Db 2890 gttcagagataataacccaacgggtcccgaaaccttataaacccttgagccctccgcct 2949
QY 2940 gaacggaactgtgtacacagatattgacttaagaatgctctctcctcgagatlaac 2999
Db 2950 gaacggaactgtgtacacagatattgacttaagaatgctctctcctcgagatlaac 3009
QY 3000 cccactagccaacacacttctgcttggaatgagagatcccaagtcacgggaagaacggg 3059
Db 3010 cccactagccaacacacttctgcttggaatgagagatcccaagtcacgggaagaacggg 3069
QY 3060 cagctcacttgaccccgactgcccgaagggttcaagaactcccccagacactcttgagaa 3119
Db 3070 cagctcacttgaccccgactgcccgaagggttcaagaactcccccagacactcttgagaa 3129
QY 3120 gccctacacaggaagcctggaacacttaagatccaacacccctcaagtgacccctccag 3179
Db 3130 gccctacacaggaagcctggaacacttaagatccaacacccctcaagtgacccctccag 3189

```

```

QY 3180 tacgtgagatgacctgctcttgccggagagcccaacacagagactcttagaaggtacgaag 3239
Db 3190 tacgtgagatgacctgctcttgccggagagcccaacacagagactcttagaaggtacgaag 3249
QY 3240 gcaactcgtctggaattctctgacttaggctcacaagagcctctgctaagaagccagat 3299
Db 3250 gcaactcgtctggaattctctgacttaggctcacaagagcctctgctaagaagccagat 3309
QY 3300 tgcagagagaggttaacatacttggtgtacagtttgcggggcggaagagtgctgaag 3359
Db 3310 tgcagagagaggttaacatacttggtgtacagtttgcggggcggaagagtgctgaag 3369
QY 3360 gagcagcgaagaagaactgtagtcagatccggcccaacacacagccaagaatgaga 3419
Db 3370 gagcagcgaagaagaactgtagtcagatccggcccaacacacagccaagaatgaga 3429
QY 3420 gagttttgggagacagctggaatttgcaagctgtgacccgggttggaacctaga 3479
Db 3430 gagttttgggagacagctggaatttgcaagctgtgacccgggttggaacctaga 3489
QY 3480 gccccactctacccgcctcaacccaagaagaagggg 3514
Db 3490 gccccactctacccgcctcaacccaagaagaagggg 3524

```

RESULT 4
AAT74883 standard; cDNA: 7393 BP.
AAT74883;
AC AAT74883;
XX
DT 09-FEB-1998 (first entry)
DE
XX Porcine retrovirus cDNA (defective).
XX
KW Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW xenotransplantation; infectious; provirus; organ transplant; donor;
KW activated virus; PCR; ss.
XX
OS Porcine retrovirus.

FT	Key	Location/Qualifiers
FT	CDS	598..2172
FT		/*tag= a
FT	mat_peptide	598..2169
FT		/*tag= b
FT		/note= "putative GAG protein"
FT	CDS	2320..4737
FT		/*tag= c
FT		/note= "putative POL coding region (partial) as described in the specification"
FT	mat_peptide	2320..3522
FT		/*tag= d
FT		/note= "putative POL protein (partial)"
FT	mat_peptide	3516..4328
FT		/*tag= e
FT		/note= "putative POL protein (partial)"
FT	CDS	4332..4748
FT		/*tag= f
FT		4738..6725
FT		/*tag= g
FT		/note= "putative ENV coding region (partial) as described in the specification"
FT	mat_peptide	4752..6722
FT		/*tag= h
FT		/note= "ENV protein (partial)"
PN	MO9721836-A1.	
XX	19-JUN-1997.	
PD	13-DEC-1996;	96MO-US19680.


```
QY 1560 ggggtctcggggtcgcctcaagacgcccactaattggtcgaagtgaaagagtgatg 1619
Db 1570 ggggtctcggggtcgcctcaagacgcccactaattggtcgaagtgaaagagtgatg 1629
QY 1620 cagggaaccgaacgaacccctcgtgtattctcttgagaagtcctgaagacccctcgaagc 1679
Db 1630 cagggaaccgaacgaacccctcgtgtattctcttgagaagtcctgaagacccctcgaagc 1689
QY 1680 ttaccccttttgactcactcagaagggccagaagacccctcgtgtcgtcgtcact 1739
Db 1690 ttaccccttttgactcactcagaagggccagaagacccctcgtgtcgtcgtcact 1749
QY 1740 gggagtcgcgtcgtcgtatcagaagaaactcagaagacgtgaaggggttaagaagcgt 1799
Db 1750 gggagtcgcgtcgtcgtatcagaagaaactcagaagacgtgaaggggttaagaagcgt 1809
QY 1800 gaggtaagtgatctcagtgagagggcagaagagtgatcttcaagaaggggaacagaag 1859
Db 1810 gaggtaagtgatctcagtgagagggcagaagagtgatcttcaagaaggggaacagaag 1869
QY 1860 gagaaggaacagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1919
Db 1870 gagaaggaacagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1929
QY 1920 caagagaagaatttgactaagatctcgtcgcagtggttgaaagggagacagcagggag 1979
Db 1930 caagagaagaatttgactaagatctcgtcgcagtggttgaaagggagacagcagggag 1989
QY 1980 aagaagagaagatttaagaaataatagtcagggccctcagaagacagaagacccctgggcaat 2039
Db 1990 aagaagagaagatttaagaaataatagtcagggccctcagaagacagaagacccctgggcaat 2049
QY 2040 aggaaccacacgcgaacgaagcagtggtgtatgttaagaagaagaagaagaagaagaag 2099
Db 2050 aggaaccacacgcgaacgaagcagtggtgtatgttaagaagaagaagaagaagaagaag 2109
QY 2100 aactgcccaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2159
Db 2110 aactgcccaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2169
QY 2160 taaggagagaggggttcggaacccctcgcgaagggtaactctgaaggttgagaggg 2219
Db 2170 taaggagagaggggttcggaacccctcgcgaagggtaactctgaaggttgagaggg 2229
QY 2220 caaccagttgagttcgttgatccggagcaggaactcagtcggtcgtacaaacat 2279
Db 2230 caaccagttgagttcgttgatccggagcaggaactcagtcggtcgtacaaacat 2289
QY 2280 ggaagaactaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2339
Db 2290 ggaagaactaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2349
QY 2340 tggactaccgaagaagacgttgacttggtgagtgagcgggtaaccactggtcttcgtc 2399
Db 2350 tggactaccgaagaagacgttgacttggtgagtgagcgggtaaccactggtcttcgtc 2409
QY 2400 atccctggtggtccagtaacccctctagtgagagacttactgacaaagtggagctcaa 2459
Db 2410 atccctggtggtccagtaacccctctagtgagagacttactgacaaagtggagctcaa 2469
QY 2460 attcttttgacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2519
Db 2470 attcttttgacaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 2529
QY 2520 accctcaatagatgataatcgaactatacttccccaagtaaaagcctgataagat 2579
Db 2530 accctcaatagatgataatcgaactatacttccccaagtaaaagcctgataagat 2589
QY 2580 atacagtcctggttgagcaagtttcccaagcctgggagagaacgcagagatgggtttg 2639
Db 2590 atacagtcctggttgagcaagtttcccaagcctgggagagaacgcagagatgggtttg 2649
QY 2649 gcaaaagcaagtttccccaagttatctcaactgaagcagtgctacacagatcatgctc 2699
```

```
Db 2650 gcaaaagcaagtttccccaagttatctcaactgaagcagtgctacacagatcatgctc 2709
QY 2700 agacagtaaccccttgatgagaagcgtcgaagaagaatttggccgacgttcaagaagat 2759
Db 2710 agacagtaaccccttgatgagaagcgtcgaagaagaatttggccgacgttcaagaagat 2769
QY 2760 atcccaagggcatcctagttcctcgtcccaatcccttggaatactccctgctacaggt 2819
Db 2770 atcccaagggcatcctagttcctcgtcccaatcccttggaatactccctgctacaggt 2829
QY 2820 aggaagccttggaacccaatgatatcgaacagtcacaggaacttgagaaggtcaataaag 2879
Db 2830 aggaagccttggaacccaatgatatcgaacagtcacaggaacttgagaaggtcaataaag 2889
QY 2880 gtgcagagacataacccaacagtcgccgaaccccttaaacctcttgagcgcctccgct 2949
Db 2890 gtgcagagacataacccaacagtcgccgaaccccttaaacctcttgagcgcctccgct 2959
QY 2940 gaaacgaactgatacagaatgagcttaaaagatgctctcctcgtcgtagatcac 2999
Db 2950 gaaacgaactgatacagaatgagcttaaaagatgctctcctcgtcgtagatcac 3009
QY 3000 cccaactagccaacacatttttgctcgtcgaatgagaagatccaggtacggagaagaacgg 3059
Db 3010 cccaactagccaacacatttttgctcgtcgaatgagaagatccaggtacggagaagaacgg 3069
QY 3060 cagctacactggaacccagatcgcgaaggttcagaagaatccccaacacatcttgagaa 3119
Db 3070 cagctacactggaacccagatcgcgaaggttcagaagaatccccaacacatcttgagaa 3129
QY 3120 gccctacagaaggacacttgccaacttcaggaatcccaacacatccctcctccag 3179
Db 3130 gccctacagaaggacacttgccaacttcaggaatcccaacacatccctcctccag 3189
QY 3180 taagtgagtgacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 3239
Db 3190 taagtgagtgacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 3249
QY 3240 gcaactacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 3299
Db 3250 gcaactacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 3309
QY 3300 tgcagagagagaggaataacttggtgagtcaggttcggtgagcgggacagtcgtcgtcgt 3359
Db 3310 tgcagagagagaggaataacttggtgagtcaggttcggtgagcgggacagtcgtcgtcgt 3369
QY 3360 gaggcagagagaagaacttgatccagataccggcccaacacagaagaagaagaag 3419
Db 3370 gaggcagagagaagaacttgatccagataccggcccaacacagaagaagaagaag 3429
QY 3420 gaatttttggaagacgttgatcttgacagacttgatccggggttcgacactagca 3479
Db 3430 gaatttttggaagacgttgatcttgacagacttgatccggggttcgacactagca 3489
QY 3480 gcccaactacccgttaacaaagaagaagaagaagaagaagaagaagaagaagaagaaga 3514
Db 3490 gcccaactacccgttaacaaagaagaagaagaagaagaagaagaagaagaagaagaaga 3524
```

```
RESULT 5
AAV09698
ID AAV09698 standard; DNA: 3320 BP.
AC
XX AAV09698;
XX
XX
XX 19-MAY-1998 (first entry)
XX
XX Porcine retrovirus pol and env DNA.
XX
XX Porcine retrovirus: PoEV; POL protein; ENV protein; vaccine;
XX diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
XX
```

OS Porcine retrovirus.
 XX Key Location/Qualifiers
 FH CDS 23..2793
 FT /*tag= a
 FT /product= pol protein
 FT /note= "polymerase protein"
 FT 2642..3287
 CDS
 FT /*tag= D
 FT /product= ENV protein
 FT /note= "envelope protein"
 XX
 XX W09740167-A1.
 PD 30-OCT-1997.
 XX
 XX 18-APR-1997; 97WO-GB01087.
 PF
 XX 10-FEB-1997; 97GB-0002668.
 PR 19-APR-1996; 96GB-0008164.
 XX
 XX (IMUT-) IMUTRAN LTD.
 PA (OONE-) O-ONE BIOTECH LTD.
 XX
 PI Galbraith DN, Haworth C, Lees GM, Smith KT;
 DR WPI: 1997-535851/49.
 XX
 XX Polynucleotide encoding porcine retrovirus expression product -
 PT useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 XX
 PS Claim 4: Fig 1: 69pp: English.
 XX
 CC This DNA sequence encodes the porcine retrovirus (PoEV) polymerase (POL)
 CC and envelope (ENV) proteins. These proteins can be used to develop viral
 CC vaccines, antisense nucleic acids, ribozymes and other antiviral agents.
 CC They can also be used in xeno-transplantation technology and as
 CC diagnostic tools.
 XX
 SQ Sequence 3320 BP; 922 A; 844 C; 832 G; 722 T; 0 other:
 Query Match 32.7%; Score 2682; DB 18; Length 3320;
 Best Local Similarity 100.0%; Freq. No. 0;
 Matches 2682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 381 ttgcggggcgagcagatgctgagcggagcaggaagaaactgagtcagatacc 440
 QY 3392 ggcaccaaccaagccaagaatgagagattttgggagacagtgtatttcagact 3451
 Db 441 ggcaccaaccaagccaagaatgagagattttgggagacagtgtatttcagact 500
 QY 3452 gtgagatccgggtttgagacacttaagcagccactctaacccgtlaacccaagaagaag 3511
 Db 501 gtgagatccgggtttgagacacttaagcagccactctaacccgtlaacccaagaagaag 560
 QY 3512 ggaattctcctgggtcttcgaagcagcagagagatttggctatataaagaagcctgt 3571
 Db 561 ggaattctcctgggtcttcgaagcagcagagagatttggctatataaagaagcctgt 620
 QY 3572 gagcgacactgtcttgccctccctgagcagtaaaccttaaccccttaattatgtatga 3631
 Db 621 gagcgacactgtcttgccctccctgagcagtaaaccttaaccccttaattatgtatga 680
 QY 3632 gctlaaggagtagcccgagagatttaaccacaaaccttaggacatttggagagactgt 3691
 Db 681 gctlaaggagtagcccgagagatttlaaccacaaaccttaggacatttggagagactgt 740
 QY 3692 tgactaactgtcaagaagagcttggatcctgtgagcagtggttgccgtatgtctgaagc 3751
 Db 741 tgactaactgtcaagaagagcttggatcctgtgagcagtggttgccgtatgtctgaagc 800
 QY 3752 tatcgacgtgtgacatactgtgtcaagagcgtgacaaattgactttggagagaatat 3811
 Db 801 tatcgacgtgtgacatactgtgtcaagagcgtgacaaattgactttggagagaatat 860
 QY 3812 aactgtlaatagcccccatactgtgtgagagacatgttcggcagccccaagcagatgat 3871
 Db 861 aactgtlaatagcccccatactgtgtgagagacatgttcggcagccccaagcagatgat 920
 QY 3872 gaccacgcccagatgacccaactatcaaaagcgtcttcacagaagaggttcatttcgc 3931
 Db 921 gaccacgcccagatgacccaactatcaaaagcgtcttcacagaagaggttcatttcgc 980
 QY 3932 tccacacgcccgtctcaaaccccgccactcttgcctgaagagactgtgaacagatgac 3991
 Db 981 tccacacgcccgtctcaaaccccgccactcttgcctgaagagactgtgaacagatgac 1040
 QY 3992 tcatgattgcatcaactatgtattgagagagctgggttcgcgaagagccttaacgaat 4051
 Db 1041 tcatgattgcatcaactatgtattgagagagctgggttcgcgaagagccttaacgaat 1100
 QY 4052 accgtgactggaagagtgcttaacctggttcaactgaacggaagcagatgtgtgaaag 4111
 Db 1101 accgtgactggaagagtgcttaacctggttcaactgaacggaagcagatgtgtgaaag 1160
 QY 4112 taagaagatgtgtggggcgagcagtggtgaacgagccgacagatctgggcaagacct 4171
 Db 1161 taagaagatgtgtggggcgagcagtggtgaacgagccgacagatctgggcaagacct 1220
 QY 4172 gccgaaggaactcaacgcaaaagcgtgactgaactatgagcctcaacgaagctttggcgt 4231
 Db 1221 gccgaaggaactcaacgcaaaagcgtgactgaactatgagcctcaacgaagctttggcgt 1280
 QY 4232 ggcgaaggaagaaatccataaacttatacggacagagatgagcctttgagctgaca 4291
 Db 1281 ggcgaaggaagaaatccataaacttatacggacagagatgagcctttgagctgaca 1340
 QY 4292 cgtacaaggagcatalctataaacaaggggttgcttaactcagcagggaggaataaa 1400
 Db 1341 cgtacaaggagcatalctataaacaaggggttgcttaactcagcagggaggaataaa 1460
 QY 4352 gaacaaagagaaatctcaagcctattagaagccttaattgccaagaagcctagat 4411
 Db 1401 gaacaaagagaaatctcaagcctattagaagccttaattgccaagaagcctagat 1460
 QY 4412 tatcaactgtctgtgacatcgaagaagccaagatctcatatctagaaggaacagatggc 4471
 Db 1461 tatcaactgtctgtgacatcgaagaagccaagatctcatatctagaaggaacagatggc 1520

QY	4472	ggacgggtgtgcgaagcagcagcccaagcgtgttaaaccctctgcctatataagaagcc	4531
Db	1521	tgaccgggtctgcgaagcagcccaagcgtgttaaaccctctgcctatataagaagcc	1581
QY	4532	caaaagcccccgaagcccaagacagacagacaccccttagaagctgycgaagagataaaaaagat	4591
Db	1581	caaaagcccccgaagcccaagacagacagacaccccttagaagctgycgaagagataaaaaagat	1640
QY	4592	agacccagttctctcgaactccggaggggagccctgtcatacctatattggaagaaatcct	4651
Db	1641	agacccagttctctcgaactccggaggggagccctgtcatacctatattggaagaaatcct	1700
QY	4652	gcccccaagaagaaggttagaataatgtccaaagaataatcgtctaaaccctaaggaaac	4711
Db	1701	gcccccaagaagaaggttagaataatgtccaaagaataatcgtctaaaccctaaggaaac	1760
QY	4712	taaacacctgcagcaggttggttcagaacatacccttatcatgtctcgaagctacagaagt	4771
Db	1761	taaacacctgcagcaggttggttcagaacatacccttatcatgtctcgaagctacagaagt	1820
QY	4772	ggctgactcgtgtgtcacaacatttgtgtccctgcgcagctgtgttaaagtcatcccttcag	4831
Db	1821	ggctgactcgtgtgtcacaacatttgtgtccctgcgcagctgtgttaaagtcatcccttcag	1880
QY	4832	aataccctccaggaagaagacacccaaggggaagccaccggcgctcaactcagggaagtgtact	4891
Db	1881	aataccctccaggaagaagacacccaaggggaagccaccggcgctcaactcagggaagtgtact	1940
QY	4892	cacttaaggtaaaagccggttaataatccggaacaaatactatctgtgtttgttgaacacctt	4951
Db	1941	cacttaaggtaaaagccggttaataatacggaaacaaatactatctgtgtttgttgaacacctt	2000
QY	4952	ttcagagatgggttaagagctatacctactaaagaagacacttcaaccgttggttgctaaaga	5011
Db	2001	ttcagagatgggttaagagctatacctactaaagaagacacttcaaccgttggttgctaaaga	2060
QY	5012	aatactgagagaataattttccaagatttggataactaaagtgaataatgaaggttcagacatatg	5071
Db	2061	aatactgagagaataattttccaagatttggataactaaagtgaataatgaaggttcagacatatg	2120
QY	5072	ttcagacttcgtttgccccaggtaaagtcacagagactcgtgcacaagatttgggattgtattgaa	5131
Db	2121	ttcagacttcgtttgccccaggtaaagtcacagagactcgtgcacaagatttgggattgtattgaa	2180
QY	5132	actgcattgtgcatacagaacccccaaaagctccaggaacaggtagaagagatgaatagaacat	5191
Db	2181	actgcattgtgcatacagaacccccaaaagctccaggaacaggtagaagagatgaatagaacat	2240
QY	5192	taaaagagacccttaaccaaatgtgacccaagagctgtgcattaaatgtttgattgagctctcct	5251
Db	2241	taaaagagacccttaaccaaatgtgacccaagagctgtgcattaaatgtttgattgagctctcct	2300
QY	5252	ggcccttgggtcttttaagggtgagagaaacccccttgaaagatttggcctgtgacccctatga	5311
Db	2301	ggcccttgggtcttttaagggtgagagaaacccccttgaaagatttggcctgtgacccctatga	2360
QY	5312	attgtcctacaggggagaccccccccgcttttgcaagaaattgcctttgcacatagctgcatagt	5371
Db	2361	attgtcctacaggggagaccccccccgcttttgcaagaaattgcctttgcacatagctgcatagt	2420
QY	5372	gctgctttcccaagccttgtctctctaggtccaaagcgtctgagtgtggttgaagcagcagac	5431
Db	2421	gctgctttcccaagccttgtctctctaggtccaaagcgtctgagtgtggttgaagcagcagac	2480
QY	5432	gtggaagacagctcccggaagcgtcactcaaggggagagacttgcaagtctccacatcgtctcca	5491
Db	2481	gtggaagacagctcccggaagcgtcactcaaggggagagacttgcaagtctccacatcgtctcca	2540
QY	5492	agtttgagatctcaagtctatgtttaagcgcacccgtgcaggaagaacctcagagactcgtgtgaa	5551
Db	2541	agtttgagatctcaagtctatgtttaagcgcacccgtgcaggaagaacctcagagactcgtgtgaa	2600

Accession	Sequence	Length
OY_5552	ggagacttatctgtaactttggacacacaaagcgtctggaagtcgcgaagatccac	5611
Db_2601	ggagacttatctgtaactttggacacacaaagcgtctggaagtcgcgaagatccac	2666
OY_5612	ctgagatcatgcatccacagcttaagcgcgcgaactccga	5653
Db_2661	ctgagatcatgcatccacagcttaagcgcgcgaactccga	2702

RESULT	6
AAV09703	
ID	AAV09703 standard; DNA; 1974 BP
XX	

20-MAY-1998 (first entry)

Porcine retrovirus Raj1 clone ENV DNA.

diagnosis; xenotransplantation; prophylactic; therapeutic; ds.

OS Porcine retrovirus.

EH	Key	Location/Qualifiers
FT	CDS	1 1074

FT	CDS	Location/Qualifiers
		1..1974

```

/product= ENV protein
/note= "envelope protein"

```

PN WO9740167-A1

PD 30-OCT-1997

PF 18-APR-1997; 97WO-GB01087

PR	10-FEB-1997;	97GB-0002668.
PR	19-APR-1996;	95GB-0000154.

PR 19-APR-1996; 96GB-0008164.
XY

PA
QV
(QONE-) Q-ONE BIOTECH LTD.

Galbraith DN, Haworth C, Lees GM, Smith KT;

WPI; 1997-535851/49

P-PSDB; AAW39274.

poly nucleotide encoding porcine retrovirus expression product -
useful to develop products for use in vaccines, diagnosis and
xeno-transplantation

Claim 6; Fig 4; 69pp; English.

ins sequence encodes the porcine retrovirus envelope (ENV) protein isolated from the human cell line Raji. Such viral proteins can be used to develop viral vaccines, antisense nucleic acids, ribozymes and other antiviral agents. They can also be used in xeno-transplantation technology and as diagnostic tools.

Sequence 1974 BP; 534 A; 489 C; 483 G; 468 T; 0 other;

Query Match	18.38:	Score 1499.	DB 18.	Length 1074
-------------	--------	-------------	--------	-------------

Local Similarity	99.18	Pred. No. 0
Matches 1949; conservative	0	Mismatches

```
..... conservative 0; Mismatches 3; Indels 3; Gaps 2;
```

5698

...ccgagacgcgaatccttcaagcttcg 79

b
80 cctccatcgctgattccttactctaaacataactcccccaagccacttatcccttcctttt

QY 5759 tagacagctcgaaaccccatagaccttacccttactggtgattatagaccctgata 5818
DB 140 tagacagctcgaaaccccatagaccttacccttactggtgattatagaccctgata 199
QY 5819 cgggtgtcctgttaaaatgaactcgaaggtgtgtgtcctcctagaagcgactgtgtcctgaac 5878
DB 200 cgggtgtcctgttaaaatgaactcgaaggtgtgtgtcctcctagaagcgactgtgtcctgaac 259
QY 5879 tgcatcttcgctccgactgattatgaaccccgctgttaaaagcacactcccaactgaagc 5938
DB 260 tgcatcttcgctccgactgattatgaaccccgctgttaaaagcacactcccaactgaagc 319
QY 5939 gtagttatgggtctctatgtctgtcccgagacagagaagaagaataactgtgtgtgtcgt 5998
DB 320 gtagttatgggtctctatgtctgtcccgagacagagaagaagaataactgtgtgtgtcgt 379
QY 5999 gggaaactctctgtagagatgtagctgtcgtcaactcccaagatgtagactgtgaatggtc 6058
DB 380 gggaaactctctgtagagatgtagctgtcgtcaactcccaagatgtagactgtgaatggtc 439
QY 6059 cgatctctccacagacgggttaaaatctctcttgcattccggcccgagcaagataca 6118
DB 440 cgatctctccacagacgggttaaaatctctcttgcattccggcccgagcaagataca 499
QY 6119 aaatgataaactatataagatlaagactgtcctcccatcagacttagattatctaaaga 6178
DB 500 aaatgataaactatataagatlaagactgtcctcccatcagacttagattatctaaaga 559
QY 6179 taagtctactg--aaagaaacagagaataatctcaaaagtgtgataatggtatagct 6236
DB 560 taagtctactgaaagaaacagagaataatctcaaaagtgtgataatggtatagct 619
QY 6237 ggggaatagttttttat-tatatgacggggagagaggtgcactttaaccattcgcctta 6295
DB 620 ggggaatagttttttatataatgtgcggggagaggtgcactttaaccattcgcctta 679
QY 6296 ggaatgagacggggagacagaaacccctgtgcaatgaggacccgataaagtgtcgtgaac 6355
DB 680 ggaatgagacggggagacagaaacccctgtgcaatgaggacccgataaagtgtcgtgaac 739
QY 6356 aagggtcccccgtcctgtgagcgacacgcaactgtcgtgcccataaactcgtcgtc 6415
DB 740 aagggtcccccgtcctgtgagcgacacgcaactgtcgtgcccataaactcgtcgtc 799
QY 6416 ggcctgacataacacagcgccctagcaacagtaacactggtatgtattcctcaaacagc 6475
DB 800 ggcctgacataacacagcgccctagcaacagtaacactggtatgtattcctcaaacagc 859
QY 6476 ctagaactccacaggtgttctgtttaaagacagagacagactcttcagctcatalcag 6535
DB 860 ctagaactccacaggtgttctgtttaaagacagagacagactcttcagctcatalcag 919
QY 6536 gaagcttccaaagcatalcaactccacgacactgtatgccaactctctctgtgtgtc 6595
DB 920 gaagcttccaaagcatalcaactccacgacactgtatgccaactctctctgtgtgtc 979
QY 6596 tatcctcaagggtcctcttattatgagggatggtctaaagaagaataatcaatgtgaca 6655
DB 980 tatcctcaagggtcctcttattatgagggatggtctaaagaagaataatcaatgtgaca 1039
QY 6656 aagagcataaataatcaatgtgacatggtgtgtccgaaataaagcttaacctgaagtt 6715
DB 1040 aagagcataaataatcaatgtgacatggtgtgtccgaaataaagcttaacctgaagtt 1099
QY 6716 ccgggaagggagacatgacataggaagaagctcccccaccccaacactgtgtatagta 6775
DB 1100 ccgggaagggagacatgacataggaagaagctcccccaccccaacactgtgtatagta 1159
QY 6776 ctgtgtttatgagcagcgctcagaataacagtaattagtaactggttataacaggtgt 6835
DB 1160 ctgtgtttatgagcagcgctcagaataacagtaattagtaactggttataacaggtgt 1219

QY 6836 gggacatgcaactactggtgttaacccccgtgttccacactcagcttcaaccatccaag 6895
DB 1220 gggacatgcaactactggtgttaacccccgtgttccacactcagcttcaaccatccaag 1279
QY 6896 attctgtgtcattgtgtccaaatcgtcccccaggtgttactacatcctgaggatgtc 6955
DB 1280 attgtgtgtcattgtgtccaaatcgtcccccaggtgttactacatcctgaggatgtc 1339
QY 6956 ttagatgaatagactgtgttaacccgacccaagaagaacccgtatcccttaacctag 7015
DB 1340 ttagatgaatagactgtgttaacccgacccaagaagaacccgtatcccttaacctag 1399
QY 7016 ctgtatgtctcagatattaggaacggtgtgtgtgtggaacagggacgtgcccgtatca 7075
DB 1400 ctgtatgtctcagatattaggaacggtgtgtgtgtggaacagggacgtgcccgtatca 1459
QY 7076 caagacacacagccttagagaaagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7135
DB 1460 caagacacacagccttagagaaagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1519
QY 7136 tccgagccttaaggaatgt 7195
DB 1520 tccgagccttaaggaatgt 1579
QY 7196 tgtgtctacagaacccggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7255
DB 1580 tgtgtctacagaacccggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1639
QY 7256 cagccttaagaagaatgt 7315
DB 1640 cagccttaagaagaatgt 1699
QY 7316 tgaacaagcttagaanaaaagttagagagcgctgaagaagaagaagagctgacaggt 7375
DB 1700 tgaacaagcttagaanaaaagttagagagcgctgaagaagaagaagagctgacaggt 1759
QY 7376 ggtttgaaggt 7435
DB 1760 ggtttgaaggt 1819
QY 7436 ggcctcctagtagtctgt 7495
DB 1820 ggcctcctagtagtctgt 1879
QY 7496 ttagccttgttagagaacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7555
DB 1880 ttagccttgttagagaacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1939
QY 7556 aagcgtctgtgacgaagaagaactgacctag 7590
DB 1940 aagcgtctgtgacgaagaagaactgacctag 1974

RESULT 7
ID AAV82749 standard; DNA: 3482 BP.
AC AAV82749:
XX 25-FEB-1999 (first entry)
DE Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
XX Pig endogenous retrovirus; PERV-A: envelope protein; PERV-B: subtype;
KW probe: detection; retrovirus; human tissue; xenotransplant;
KW primary porcine tissue; human cell line; porcine cell line; ss.
XX Pig endogenous retrovirus.
OS
XX
FH Key Location/Qualifiers
FT CDS 911..2884
FT /*tag= a
FT /product= envelope_protein

XX MO9853104-A2.
 XX 26-NOV-1998.
 XX 18-MAY-1998; 98WO-GB01428.
 XX 16-MAY-1997; 97GB-0010154.
 XX (MEDI-) MEDICAL RES COUNCIL.
 XX Stoye JP, Weiss RA;
 XX WPI: 1999-045324/04.
 XX P-PSDB: AAW85452.
 XX Newly isolated nucleic acid probe capable of hybridizing to either
 XX P1 PERV-A or PERV-B env gene - useful in the detection of
 XX P2 retroviruses, and their subtypes, in a sample of porcine/human
 XX tissue
 XX Claim 3: Page 21-23; 36pp; English.
 XX The present sequence encodes a pig endogenous retrovirus (PERV)-B
 XX envelope protein. PERV exists in two different subtypes, PERV-A and
 XX PERV-B. The differences are reflected in sequence divergence in the
 XX (env) genes of PERV-A and PERV-B. The probes and primers are used in
 XX a method to detect retroviruses in a sample of porcine/human tissue,
 XX particularly primary porcine tissue and human cell lines that have been
 XX cultivated in the presence of a porcine cell line, or human tissue from
 XX one of the PERV env genes can also be determined.
 XX Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;

Query Match 17.9%; Score 1466; DB 20; Length 3482;
 Best Local Similarity 99.38; Pred. No. 0;
 Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;

QY 5639 ggcgcacccctccgactcgggtgagagccgaagaagctgagaatcccttaagcttg 5698
 DB 930 ggcgcacccctccgactcgggtgagagccgaagaagctgagaatcccttaagcttg 5698
 QY 5699 cctccatcgctggttcttacttaacaaataactcccaagcagtagtaagcctta 5758
 DB 990 cctccatcgctggttcttacttaacaaataactcccaagcagtagtaagcctta 5758
 QY 5759 taagaagctcgaacccccaatagaccccttaacccctgctgctgattatgaaccc 5818
 DB 1050 taagaagctcgaacccccaatagaccccttaacccctgctgctgattatgaaccc 5818
 QY 5819 cgggtgtcaactgttaataagcaactcgaagtggtctcctcgaagcactggtgcttg 5878
 DB 1110 cgggtgtcaactgttaataagcaactcgaagtggtctcctcgaagcactggtgcttg 5878
 QY 5879 tgcattctgcgcctcgatgattaaacccgcgtgttaaaagcaacactcccaactagctc 5938
 DB 1170 tgcattctgcgcctcgatgattaaacccgcgtgttaaaagcaacactcccaactagctc 5938
 QY 5939 gtagtatagtgtctatctgctccagcagacagaaagaaataactgtggtgctctg 5998
 DB 1230 gtagtatagtgtctatctgctccagcagacagaaagaaataactgtggtgctctg 5998
 QY 5999 ggggaatcctctgttagagagatgagctgctcactccaagatgagagactggaatggc 6058
 DB 1290 ggggaatcctctgttagagagatgagctgctcactccaagatgagagactggaatggc 6058
 QY 6059 cgatctctccagagacgggttaaattctctcttgcactcccgccggcggaagtaaca 6118
 DB 1350 cgatctctccagagacgggttaaattctctcttgcactcccgccggcggaagtaaca 6118

QY 6119 aaatgatgaactatataaagaagctgctcccatcaagacttagatctataaaga 6178
 DB 1410 aagtgatgaactatataaagaagctgctcccatcaagacttagatctataaaga 6178
 QY 6179 taagtctacg--aaggaacacaggaataatcacaagtgataatgataagct 6236
 DB 1470 taagtctacg--aaggaacacaggaataatcacaagtgataatgataagct 6236
 QY 6237 ggggaatagttttat-tatagcgggggagcaggggtccacttaacacttcgctta 6295
 DB 1530 ggggaatagttttat-tatagcgggggagcaggggtccacttaacacttcgctta 6295
 QY 6296 ggataagagcgggagacagaacccctgtgcaatggaccagataaagtaactgctgac 6355
 DB 1590 ggataagagcgggagacagaacccctgtgcaatggaccagataaagtaactgctgac 6355
 QY 6356 aaggggcccccggccggggagccacacatacttcggtgcccataacttcgctgac 6415
 DB 1650 aaggggcccccggccggggagccacacatacttcggtgcccataacttcgctgac 6415
 QY 6416 ggcctgacataacacagccgcttagcaacagatgacactggtatcttcccaacac 6475
 DB 1710 ggcctgacataacacagccgcttagcaacagatgacactggtatcttcccaacac 6475
 QY 6476 ctagaacaccccaagtggtctctgttaagaacagagacactctcaagctcagcag 6535
 DB 1770 ctagaacaccccaagtggtctctgttaagaacagagacactctcaagctcagcag 6535
 QY 6536 gagcttccaaagcctcactccacacccctgtgcaactctctgttgcctgtgc 6595
 DB 1830 gagcttccaaagcctcactccacacccctgtgcaactctctgttgcctgtgc 6595
 QY 6596 tatctcagggcctccttattatagggagtggtcctaaagaagaataatcaagtgacca 6655
 DB 1890 tatctcagggcctccttattatagggagtggtcctaaagaagaataatcaagtgacca 6655
 QY 6656 aagagacataagaatcaatgatacgtggtggtcccaataagcttaccctcgaagttt 6715
 DB 1950 aagagacataagaatcaatgatacgtggtggtggtcccaataagcttaccctcgaagttt 6715
 QY 6716 ccgggaagggagcgtatgataagaaagctcccatcccaacacacttgcataagta 6775
 DB 2010 ccgggaagggagcgtatgataagaaagctcccatcccaacacacttgcataagta 6775
 QY 6776 ctgtgttatacagcagcctcagaataatcagattatgactggttataacagtgct 6835
 DB 2070 ctgtgttatacagcagcctcagaataatcagattatgactggttataacagtgct 6835
 QY 6836 gggcatgcaactctggttaacccctgtgttccacactcagcttcaacacacacaaag 6895
 DB 2130 gggcatgcaactctggttaacccctgtgttccacactcagcttcaacacacacaaag 6895
 QY 6896 attctgtgcaatggtgcgaatcgttcccccgggtgtactacatctgaggaagtgctcc 6955
 DB 2190 attctgtgcaatggtgcgaatcgttcccccgggtgtactacatctgaggaagtgctcc 6955
 QY 6956 ttgataataatgatactcgtataacacgaacaaagaagaaacccgtatccctcctag 7015
 DB 2250 ttgataataatgatactcgtataacacgaacaaagaagaaacccgtatccctcctag 7015
 QY 7016 ctgataatgctgataagggagcgcgtgtggtcgttagaagcgggagcgtgctcgtatca 7075
 DB 2310 ctgataatgctgataagggagcgcgtgtggtcgttagaagcgggagcgtgctcgtatca 7075
 QY 7076 caggacccacagacgtataagaagaagactgtgtgctcactatcggtcgtatcaagaagatc 7135
 DB 2370 caggacccacagacgtataagaagaagactgtgtgctcactatcggtcgtatcaagaagatc 7135
 QY 7136 tccagccttaagaagatcgtttagcaacttagaagaagctccgactcttcttgcgaag 7195
 DB 2430 tccagccttaagaagatcgtttagcaacttagaagaagctccgactcttcttgcgaag 7195
 QY 7196 tggttctacagaacccgagggatagatctgcttcttcaagaagaagtggttatgct 7255

Db 2490 tggctctacagacccggaaggagattagatcgtgttctcaagagaagtgtgtatgtg 2549
 QY 7256 cagccttaaaagaagaatgttctctatagatagatcaactcagagacatcagagactcca 7315
 Db 2550 cagccttaaaagaagaatgttctctatagatagatcaactcagagacatcagagactcca 2609
 QY 7316 tgaacaagcttagaaaaaattagagagagcgtcgaaaggaaagagagcttagcagagaggt 7375
 Db 2610 ttagcaagcttagaagaagtttagagagagcgtcgaaaggaaagagagcttagcagagaggt 2669
 QY 7376 gggttgaagatgttcaacaaggtctcccttgatgagacccacccgtcttctgtctgaag 7435
 Db 2670 gggttgaagatgttcaacaaggtctcccttgatgagacccacccgtcttctgtctgaag 2729
 QY 7436 ggcctctagtagtctgtctctgttacttaagttgggcttggccttgaatlaagtgtg 7495
 Db 2730 gacccttagtagtctgtctctgttacttaagttgggcttggccttgaatlaagtgtg 2789
 QY 7496 ttgctcttgttagagaagcaggtgagtgagtgccagatcagatgagtagtaggcaacagtagc 7555
 Db 2790 ttgctcttgttagagaagcaggtgagtgagtgccagatcagatgagtagtaggcaacagtagc 2849
 QY 7556 aagcgtcttagcagcagaagaaacttagacttagccttccagatctctaagatlaaact 7615
 Db 2850 aagcgtcttagcagcagaagaaacttagacttagccttccagatctctaagatlaaact 2909
 QY 7616 attacaagaacaagaagtgagggaattgaagaattgaataatgaacttaaccctccagac 7675
 Db 2910 attacaagaacaagaagtgagggaattgaagaattgaataatgaacttaaccctccagac 2969
 QY 7676 ccaagaagttataaagaagccttaaaabgcccccaatlaacagaccctgtgtgtgcaggt 7735
 Db 2970 ccaagaagttataaagaagccttaaaabgcccccaatlaacagaccctgtgtgtgcaggt 3029
 QY 7736 aatatagtagaagtgacacacttctctatgttccagaggtctctctctgtgacctgaagtag 7795
 Db 3030 aatatagtagaagtgacacacttctctatgttccagaggtctctctctgtgacctgaagtag 3089
 QY 7796 ataaagagaagaatgagttgactatcgtatctgattctgtlaaaacttagctggacacat 7855
 Db 3090 ataaagagaagaatgagttgactatcgtatctgattctgtlaaaacttagctggacacat 3149
 QY 7856 agaagaattgatatcacatcttagacagccttagtgacctatctcaactgaactcttcaactc 7915
 Db 3150 agaagaattgatatcacatcttagacagccttagtgacctatctcaactgaactcttcaactc 3209
 QY 7916 tggccagagagccagacagatgagacccctcgagagctatcttaaatgattgtccacgg 7975
 Db 3210 tggccagagagccagacagatgagacccctcgagagctatcttaaatgattgtccacgg 3269
 QY 7976 agcgcgggtctcgtatattttaaatgattgttccatgagagcgcgctc-tcgatatt 8034
 Db 3270 agcgcgggtctcgtatattttaaatgattgttccatgagagcgcgctc-tcgatatt 8094
 QY 8035 taaatgattgttgttgaagcagagcttgttgaagcccatlaaaagcgttcccgat 8094
 Db 3330 taaatgattgttgttgaagcagagcttgttgaagcccatlaaaagcgttcccgat 3389
 QY 8095 tccgcactcgggcgagcagctctctacccctcggtgtgtgaagcgttgggccccagcg 8154
 Db 3390 tccgcactcgggcgagcagctctctacccctcggtgtgtgaagcgttgggccccagcg 3449
 QY 8155 gcttggataaaatctcttgcgtgttgcac 8187
 Db 3450 gcttggataaaatctcttgcgtgttgcac 3482

RESULT 8
 AAC67020
 ID AAC67020 standard; DNA: 4402 BP.
 XX
 AC AAC67020;

XX 27-MAR-2001 (first entry)
 DT
 XX PERV env protein coding sequence SEQ ID NO: 20.
 XX
 XX Xenoctransplantation; infectious agent; vaccine; ds.
 XX
 XX Porcine endogenous retrovirus.
 OS
 MO200071726-AL.
 PN
 PN 30-NOV-2000.
 PD
 PF 24-MAY-2000; 2000MO-US14296.
 PF
 PR 24-MAY-1999; 99US-0135631.
 PR
 PA (MAYO-) MAYO MEDICAL VENTURES.
 PA
 PI Federpsiel MJ;
 PI
 DR WPI: 2001-032041/04.
 DR
 XX Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PS
 PS Claim 16; Page 105-106; 144pp: English.
 XX
 XX The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 CC
 SQ Sequence 4402 BP; 1259 A; 1085 C; 1111 G; 947 T; 0 other;
 SQ
 Query Match 10.3%; Score 848; DB 22; Length 4402;
 Best Local Similarity 98.8%; Pred. NO. 0;
 Matches 2198; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 2982 tctctgctgagatttaacccctcagtagcccaacatttctgctctgaatgagagatcca 3041
 Db 1043 tctctgctgagatttaacccctcagtagcccaacatttctgctctgaatgagagatcca 1102
 QY 3042 ggtacgggaagaacccggcagctcagcttagaccgagctgcccagaaggttcaagaactcc 3101
 Db 1103 ggtacgggaagaacccggcagctcagcttagaccgagctgcccagaaggttcaagaactcc 1162
 QY 3102 ccgacacatcttgaagaaagccctacacagagacccctggccaacttcaagatccaacacct 3161
 Db 1163 ccgacacatcttgaagaaagccctacacagagacccctggccaacttcaagatccaacacct 1222
 QY 3162 caggtgacccctctccagtagtcgtagtgcactgtctctgctggagagcccaacaagagac 3221
 Db 1223 caggtgacccctctccagtagtcgtagtgcactgtctctgctggagagcccaacaagagac 1282
 QY 3222 tgcctagaagtagaagaagcagctacactgcttgaatttctgaactgaagctacagagcctct 3281
 Db 1283 tgcctagaagtagaagaagcagctacactgcttgaatttctgaactgaagctacagagcctct 1342
 QY 3282 gctagaagaagccagagatttgcagagagaggttaacataacttgggtgtagagtttgcggg 3341
 Db 1343 gctagaagaagccagagatttgcagagagaggttaacataacttgggtgtagagtttgcggg 1402
 QY 3342 gggcagcgaatgctgacgagagcagcgaagaagaactgagttcaagtagacggcccaacc 3401
 Db 1403 gggcagcgaatgctgacgagagcagcgaagaagaactgagttcaagtagacggcccaacc 1462
 QY 3402 acagcacaagaatgagagagtttcttgggacagctgagatttgcagactgtgagatccg 3461

Db 1463 aacgccaacaaagtgaagaggttttttgggaacgtgtgatttgcagactgtgatccg 1522
 QY 3462 gggtttgcgaccttagcagcccaactactacccgttaacccaagaagaagggtatcttc 3521
 Db 1523 gggtttgcgaccttagcagcccaactactacccgttaacccaagaagaagggtatcttc 3521
 QY 3522 tgggtctctggcccggaagagcatttgaatgctacaaaaggccctgtcgagcagact 3581
 Db 1583 tgggtctctggcccggaagagcatttgaatgctacaaaaggccctgtcgagcagact 3581
 QY 3582 gctctggccctctggcccggaagagcatttgaatgctacaaaaggccctgtcgagcagact 3641
 Db 1643 gctctggccctctggcccggaagagcatttgaatgctacaaaaggccctgtcgagcagact 3641
 QY 3642 gtagcccggaagaggttttaacccaaccccttagagacatttgagagagctgtgacactg 3701
 Db 1703 gtagcccggaagaggttttaacccaaccccttagagacatttgagagagctgtgacactg 3701
 QY 3702 tcaagaagcttgaatcccttagagacatttgagagagctgtgacactg 3761
 Db 1763 tcaagaagcttgaatcccttagagacatttgagagagctgtgacactg 3761
 QY 3762 gtggccatctgttcaagagcgttgacaaattgacttttgagacagataataactgtaata 3821
 Db 1823 gtggccatctgttcaagagcgttgacaaattgacttttgagacagataataactgtaata 3821
 QY 3822 gcccccacatgttgagagacatcgttgcgcagcccccagacagatgagatgacacagcc 3881
 Db 1883 gcccccacatgttgagagacatcgttgcgcagcccccagacagatgagatgacacagcc 3881
 QY 3882 cgcataacccaatcaaaagcctgtcttccataagagaggttgaacttgcctccacagcc 3941
 Db 1943 cgcataacccaatcaaaagcctgtcttccataagagaggttgaacttgcctccacagcc 3941
 QY 3942 gctcctaaccgacacacttcttgcctgaagaagactgtgaacagagctacatgactgtgc 4001
 Db 4003 gctcctaaccgacacacttcttgcctgaagaagactgtgaacagagctacatgactgtgc 4001
 QY 4002 catcaactatgtatgagagagcgttgcgcagacagcccttgaagacataacgctgact 4061
 Db 2063 catcaactatgtatgagagagcgttgcgcagacagcccttgaagacataacgctgact 4061
 QY 4062 ggaagaagtgtaactgttcaactgaacgagacatgtgtgagaggttaagagagatg 4121
 Db 2123 ggaagaagtgtaactgttcaactgaacgagacatgtgtgagaggttaagagagatg 4121
 QY 4122 gctggggcggcagtggtgagcggagcccgacagatcgtggccagcagcctgtccggaagga 4181
 Db 2183 gctggggcggcagtggtgagcggagcccgacagatcgtggccagcagcctgtccggaagga 4181
 QY 4182 acttcagcgcaaaagcgtgagcttcaatgagccctacagcaagcttgcggtgcccgaagga 4241
 Db 2243 acttcagcgcaaaagcgtgagcttcaatgagccctacagcaagcttgcggtgcccgaagga 4241
 QY 4242 aaatccataaacttatacagcagcagatgatgcttgcgagctgacagctacacggg 4301
 Db 2303 aaatccataaacttatacagcagcagatgatgcttgcgagctgacagctacacggg 4301
 QY 4302 gccatctataaacaagaagggttgccttactacagcagcagcagcagcagcagcagcagc 4361
 Db 2363 gccatctataaacaagaagggttgccttactacagcagcagcagcagcagcagcagcagc 4361
 QY 4362 gaaatcttaagcctattagaagccttaacttgcacaaaaggctagctattataactgt 4421
 Db 2423 gaaatcttaagcctattagaagccttaacttgcacaaaaggctagctattataactgt 4421
 QY 4422 ccttgacatcagaagaacccaagaatcctacatactagaaggaacagatgctgcagcgggtc 4481
 Db 2483 ccttgacatcagaagaacccaagaatcctacatactagaaggaacagatgctgcagcgggtc 4481
 QY 4482 gccaaagcagcagccagcagctgttaacttgccttactaataaagaagcccaagcccca 4541
 Db 2543 gccaaagcagcagccagcagctgttaacttgccttactaataaagaagcccaagcccca 2602

QY 4542 gaaccacagcagactacaccccttagaagactggcagaagataaaaaagataagaccagctc 4601
 Db 2603 gaaccacagcagactacaccccttagaagactggcagaagataaaaaagataagaccagctc 2662
 QY 4602 tctgaagactccggagggagcctgtctatcccttactatcttgaaggaagaaatccctgcccacaa 4661
 Db 2663 tctgaagactccggagggagcctgtctatcccttactatcttgaaggaagaaatccctgcccacaa 2722
 QY 4662 gaaaggttaagatatgtcccaacagatatacctgtcttaacccacactagaacataacactg 4721
 Db 2723 gaaaggttaagatatgtcccaacagatatacctgtcttaacccacactagaacataacactg 2782
 QY 4722 cagcagcttgcctgaacatacccttactatcttgaagcctacagagtggtgctgacgc 4781
 Db 2783 cagcagcttgcctgaacatacccttactatcttgaagcctacagagtggtgctgacgc 2842
 QY 4782 gtgttcaacatgtgttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 4841
 Db 2843 gtgttcaacatgtgttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 2902
 QY 4842 ggaagaagactaaggggaagccacccagcgtctcaacttggaagtggaacttaactgagta 4901
 Db 2903 ggaagaagactaaggggaagccacccagcgtctcaacttggaagtggaacttaactgagta 2962
 QY 4902 aagcggcgttaataacggaacaaatatactattgtttttagacaaccttccagatgg 4961
 Db 2963 aagcggcgttaataacggaacaaatatactattgtttttagacaaccttccagatgg 3022
 QY 4962 gtgagagcttactactactaagaagaagacttcaacccggtgtgtgttaagaataactgag 5021
 Db 3023 gtgagagcttactactactaagaagaagacttcaacccggtgtgtgttaagaataactgag 3082
 QY 5022 gaaattttccaaagatttggaataccttaaggtaaggttgaagcaaatgttccagcttc 5081
 Db 3083 gaaattttccaaagatttggaataccttaaggtaaggttgaagcaaatgttccagcttc 3142
 QY 5082 gtgtcccgagtaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5141
 Db 3143 gtgtcccgagtaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3202
 QY 5142 gcaatcagaaccccaagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagc 5201
 Db 3203 gcaatcagaaccccaagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagc 3262
 QY 5202 cttaac 5206
 Db 3263 cttaac 3267

RESULT 9
 ID AAC67021 standard; DNM: 6076 BP.
 AC AAC67021;
 DT 27-MAR-2001 (first entry)
 DE PERV env protein coding sequence SEQ ID NO: 21.
 XX Xenoctransplantation; infectious agent; vaccine; ds.
 OS Porcine endogenous retrovirus.
 PN WO200071726-A1.
 PD 30-NOV-2000.
 XX 24-MAY-2000; 2000WO-US14296.
 PF 24-MAY-1999; 99US-0135631.
 XX 24-MAY-1999; 99US-0135631.
 PA (MAYO-) MAYO MEDICAL VENTURES.

XX Federpsiel MJ:
 XX WPI: 2001-032041/04.
 XX Inhibiting or preventing infectious agent transmission in mammalian
 XX transplant recipients, by introducing recombinant DNA comprising DNA
 XX encoding extracellular proteins of the agent into donor cells, such as
 XX swine cells -
 XX
 XX Claim 16: Page 107-109; 144pp; English.
 XX
 XX The present invention provides a method to prevent the transmission of
 XX infectious agents during xenotransplantation. This involves introducing
 XX to donor swine cells a recombinant DNA encoding a peptide fragment from
 XX the infectious agent, and then introducing these cells into the
 XX CC transplant recipient.
 XX
 XX Sequence 6076 BP: 1613 A; 1512 C; 1609 G; 1342 T; 0 other:

Query Match 9.7%; Score 797; DB 22; Length 6076;
 Best Local Similarity 98.7%; Pred. No. 1.8e-297;
 Matches 2197; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 2982 ttctgctgagattacacccactagccacacttttgccttcgaatgagagatcca 3041
 DB 3577 ttctgctgagattacacccactagccacacttttgccttcgaatgagagatcca 3636
 QY 3042 ggtacggagaagacggcgagctcaactgacacgacgcccgaagggttaagaactcc 3101
 DB 3637 ggtacggagaagacggcgagctcaactgacacgacgcccgaagggttaagaactcc 3696
 QY 3102 ccgacacactttgacgaagacccctacacagagacccggaacttcaagatccaacacct 3161
 DB 3697 ccgacacactttgacgaagacccctacacagagacccggaacttcaagatccaacacct 3756
 QY 3162 caaggtagacctcctcagtagctgtagctgcttcgctgctgagagacacacacacagac 3221
 DB 3757 caaggtagacctcctcagtagctgtagctgcttcgctgctgagagacacacacacagac 3816
 QY 3222 tgcctgaagagtagcgaagacactgctggaattgctgacactgactacagagacct 3281
 DB 3817 tgcctgaagagtagcgaagacactgctggaattgctgacactgactacagagacct 3876
 QY 3282 gctaaagaagccagagatttgcaagagagagtagaacttgggtgagcttggcgggc 3341
 DB 3877 gctaaagaagccagagatttgcaagagagagtagaacttgggtgagcttggcgggc 3936
 QY 3342 gggcagcagatgctgagcagcagcagcgaagaagaactgtagtccagaacacacacac 3401
 DB 3937 gggcagcagatgctgagcagcagcagcgaagaagaactgtagtccagaacacacacac 3996
 QY 3402 ac 3461
 DB 3997 ac 4056
 QY 3462 ggggttgcacacttaagacgcccacttaccggttaacacacacacacacacacacac 3521
 DB 4057 ggggttgcacacttaagacgcccacttaccggttaacacacacacacacacacacac 4116
 QY 3522 tgggctctgagcaccac 3581
 DB 4117 tgggctctgagcaccac 4176
 QY 3582 gctctgacccctcctgacgtagtaactaaacacttacccttaagtgggtgagcgttaag 3641
 DB 4177 gctctgacccctcctgacgtagtaactaaacacttacccttaagtgggtgagcgttaag 4236
 QY 3642 gtagcccgagaggttttaac 3701
 DB 4237 gtagcccgagaggttttaac 4296

QY 3702 tcaagaagcttgatccttgaagcagtggtgcccgtatgtctgaaggtatccgagct 3761
 DB 4297 tcaagaagcttgatccttgaagcagtggtgcccgtatgtctgaaggtatccgagct 4356
 QY 3762 gtggcattactgtcaagaagcagctgacaaatgactttgggacagatatataactgttaata 3821
 DB 4357 gtggcattactgtcaagaagcagctgacaaatgactttgggacagatatataactgttaata 4416
 QY 3822 gcccccacatgttgaggaagacatgcttcgacagcccccacacacacacacacacacacac 3881
 DB 4417 gcccccacatgttgaggaagacatgcttcgacagcccccacacacacacacacacacacac 4476
 QY 3882 cgcac 3941
 DB 4477 cgcac 4536
 QY 3942 gctctcaacccctgcacacttcctgctgaggaagacgtatgaacacacacacacacacac 4001
 DB 4537 gctctcaacccctgcacacttcctgctgaggaagacgtatgaacacacacacacacacac 4596
 QY 4002 catcaactatgattgaggaagacgtggttcgcaagacacacacacacacacacacacac 4061
 DB 4597 catcaactatgattgaggaagacgtggttcgcaagacacacacacacacacacacacac 4656
 QY 4062 ggaagaagcttaacacgttcaactgacgaagacacacacacacacacacacacacacac 4121
 DB 4657 ggaagaagcttaacacgttcaactgacgaagacacacacacacacacacacacacacac 4716
 QY 4122 gctggggcgagctgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 4181
 DB 4717 gctggggcgagctgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 4776
 QY 4182 acttcacgac 4241
 DB 4777 acttcacgac 4836
 QY 4242 aaatccataacacatattatccgacacagcagcagcagcagcagcagcagcagcagc 4301
 DB 4837 aaatccataacacatattatccgacacagcagcagcagcagcagcagcagcagcagc 4896
 QY 4302 gccatcattaaac 4361
 DB 4897 gccatcattaaac 4956
 QY 4362 gaaattcctaagcctatagaagccttaacatttgcacacacacacacacacacacacacac 4421
 DB 4957 gaaattcctaagcctatagaagccttaacatttgcacacacacacacacacacacacacac 5016
 QY 4422 ccttgagacatagaagaagcacaagaatcctatcattgacacacacacacacacacacacac 4481
 DB 5017 ccttgagacatagaagaagcacaagaatcctatcattgacacacacacacacacacacacac 5076
 QY 4482 gccaaagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4541
 DB 5077 gccaaagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5136
 QY 4542 gaaacccagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4601
 DB 5137 gaaacccagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5196
 QY 4602 tctgagactccgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4661
 DB 5197 tctgagactccgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5256
 QY 4662 gaaaggttaagaatgttccac 4721
 DB 5257 gaaaggttaagaatgttccac 5316
 QY 4722 cagcaggttgatgaagac 4781
 DB 5317 cagcaggttgatgaagac 5376
 QY 4782 gtgtgtaaacacatgtgtgctcctgacagcagcagcagcagcagcagcagcagcagc 4841

```

|||||
Db 5377 gtgtcaaaacactgtgtgtccctgcagctgtgttaattcctccagataccctca 5436
QY 4842 ggaagagagactaagggaagccaccagcgctcaactgggaagtggacttcacgtgta 4901
Db 5437 ggaagagagactaagggaagccaccagcgctcaactgggaagtggacttcacgtgta 5496
QY 4902 aagccggtataatacggaaacaatactatgtttttgttagacactttcagatgg 4961
Db 5497 aagccggtataatacggaaacaatactatgtttttgttagacactttcagatgg 5556
QY 4962 gtagaggtctactactcaagaagaagacttaacggtgtgtgttaagaataactggag 5021
Db 5557 gtagaggtctactactcaagaagaagacttaacggtgtgtgttaagaataactggag 5616
QY 5022 gaatttttccaagatttgaatacctaagtaataggtcagacacatgtccagcttc 5081
Db 5617 gaatttttccaagatttgaatacctaagtaataggtcagacacatgtccagcttc 5676
QY 5082 gttgccaggttaagtcagagactgtgcgaagatattggggaatttggaactgtcatgt 5141
Db 5677 gttgccaggttaagtcagagactgtgcgaagatattggggaatttggaactgtcatgt 5736
QY 5142 gcalacagaccccaaaagctagagacaggttagagagatgaatagaacataaagagacc 5201
Db 5737 gcalacagaccccaaaagctagagacaggttagagagatgaatagaacataaagagacc 5796
QY 5202 cttaac 5206
Db 5797 cttaac 5801

```

RESULT 10

AAC67019
ID AAC67019 standard; DNA; 7362 BP.

AC AAC67019;

DT 27-MAR-2001 (first entry)

DE PERV env protein coding sequence SEQ ID NO: 19.

KW Xenotransplantation; infectious agent; vaccine; ds.

OS Porcine endogenous retrovirus.

PN WO200071726-A1.

PD 30-NOV-2000.

PF 24-MAY-2000; 2000WO-US14296.

PR 24-MAY-1999; 99US-0135631.

PA (MAYO-) MAYO MEDICAL VENTURES.

PI Federspiel MJ;

DR WPI: 2001-032041/04.

PT Inhibiting or preventing infectious agent transmission in mammalian
transplant recipients, by introducing recombinant DNA comprising DNA
encoding extracellular proteins of the agent into donor cells, such as
swine cells -

PS Claim 16; Page 101-104; 144pp; English.

CC The present invention provides a method to prevent the transmission of
infectious agents during xenotransplantation. This involves introducing
to donor swine cells a recombinant DNA encoding a peptide fragment from
the infectious agent, and then introducing these cells into the
transplant recipient.

Sequence 7362 BP; 1997 A; 1821 C; 1881 G; 1663 T; 0 other:

Query Match 9.7%; Score 797; DB 22; Length 7362;
Best Local Similarity 98.7%; Pred. No. 1.8e-297;
Matches 2197; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```

QY 2982 ttctgcttgatattacaccaccatagccaaaccattttgcttcgaatgagaatcca 3041
Db 3642 ttctgcttgatattacaccaccatagccaaaccattttgcttcgaatgagaatcca 3701
QY 3042 ggtacgggaagaaccgggacagctcactgtgacccagctcccaagggttcagaactcc 3101
Db 3702 ggtacgggaagaaccgggacagctcactgtgacccagctcccaagggttcagaactcc 3761
QY 3102 ccgacacatctttgacgaagccctacacagagagacttgcacacttcagatccaaacct 3161
Db 3762 ccgacacatctttgacgaagccctacacagagagacttgcacacttcagatccaaacct 3821
QY 3162 caggttgacctctctccagatcgttgatgacgtctctggtcggtgaccccaacaggagc 3221
Db 3822 caggttgacctctctccagatcgttgatgacgtctctggtcggtgaccccaacaggagc 3881
QY 3222 tgcctagaaggtacgaagagcactactgtggaattgtgacttggtcagatgagacgtct 3281
Db 3882 tgcctagaaggtacgaagagcactactgtggaattgtgacttggtcagatgagacgtct 3941
QY 3282 gctaaagaagccagatattgcagagagagagtaacttgggtacagtttgagggggc 3341
Db 3942 gctaaagaagccagatattgcagagagagagtaacttgggtacagtttgagggggc 4001
QY 3342 gggcagcagatgtgctgcagagacgacggaagaactgtgagccagatccggcccaacc 3401
Db 4002 gggcagcagatgtgctgcagagacgacggaagaactgtgagccagatccggcccaacc 4061
QY 3402 acagccaaacaaagttagagagtttttggggaacgtgattttgagacgttgatcccg 3461
Db 4062 acagccaaacaaagttagagagtttttggggaacgtgattttgagacgttgatcccg 4121
QY 3462 ggggttgagacttgacgagcccaactcctacccgcgttaacaaagaagaagggaattcc 3521
Db 4122 ggggttgagacttgacgagcccaactcctacccgcgttaacaaagaagaagggaattcc 4181
QY 3522 tgggtccttgagacacaaagaagcatttgatgcttaaaaaaggccgtcgagccacct 3581
Db 4182 tgggtccttgagacacaaagaagcatttgatgcttaaaaaaggccgtcgagccacct 4241
QY 3582 gctctggccctccctgacgttaactaaaccttacccttattgtgtagcgtaaggga 3641
Db 4242 gctctggccctccctgacgttaactaaaccttacccttattgtgtagcgtaaggga 4301
QY 3642 gtagccgagagattttaaaccacacacacacacacacacacacacacacacacacac 3701
Db 4302 gtagccgagagattttaaaccacacacacacacacacacacacacacacacacacac 4361
QY 3702 tcaagaagcttgatccctgtagccagtggttggcccgtagtgaaggtacagagct 3761
Db 4362 tcaagaagcttgatccctgtagccagtggttggcccgtagtgaaggtacagagct 4421
QY 3762 gttgccaactggttcaagagcgctgacaaattgacttgggagaagataatactgtaata 3821
Db 4422 gttgccaactggttcaagagcgctgacaaattgacttgggagaagataatactgtaata 4481
QY 3822 gcccccattgcatgtagaacaactgttcgcaagcccccagacccgagatgtgcaagcc 3881
Db 4482 gcccccattgcatgtagaacaactgttcgcaagcccccagacccgagatgtgcaagcc 4541
QY 3882 cgaatgccaactatacaagcgtgcttctcaagaagaggttacttgcctccacagcc 3941
Db 4542 cgaatgccaactatacaagcgtgcttctcaagaagaggttacttgcctccacagcc 4601
QY 3942 gcttcaacctgcacctctgcttggaagagactgataacagtgactatgattgc 4001

```


Db 4602 gctctaacccctgccaactcttctgctgaagaagactgtatgaaccaagtgactatcatgttc 4661
 QY 4002 catcaactatgtatgagaagactgggttcgcaagagccttaccagatccgctgact 4061
 Db 4662 catcaactatgtatgagaagactgggttcgcaagagccttaccagatccgctgact 4771
 QY 4062 gagaagtgctaacctggttctacatcaggaagcagatctgtgtgtaagaagtaaga 4121
 Db 4722 ggaagaagtgctaacctggttctacatcaggaagcagatctgtgtgtaagaagta 4781
 QY 4122 gctggggcgagctgggtggaagggagccgacagatctgtggaagagcctgcccgaag 4181
 Db 4782 gctggggcgagctgggtggaagggagccgacagatctgtggaagagcctgcccgaag 4841
 QY 4182 actcagcgcaaaagctgagctcagcgccctcagcagccttgcgctgagcgaag 4241
 Db 4842 actcagcgcaaaagctgagctcagcgccctcagcagccttgcgctgagcgaag 4901
 QY 4242 aaatcctaacaattatatacagacagagtgatctgtggaagctgacagctaacag 4301
 Db 4902 aaatcctaacaattatatacagacagagtgatctgtggaagctgacagctaacag 4961
 QY 4302 gccatctataaacaaggggtgtgtctactctcagcgagggggaataaagaacaagag 4361
 Db 4962 gccatctataaacaaggggtgtgtctactctcagcgagggggaataaagaacaagag 5021
 QY 4362 gaaattcctaagcctttagaagcctttagccttgcgaagctgacagctaacagct 4421
 Db 5022 gaaattcctaagcctttagaagcctttagccttgcgaagctgacagctaacagct 5081
 QY 4422 cctggacatcagaagaagccaaagatctatctagaggaaggaagcagctgagcgggt 4481
 Db 5082 cctggacatcagaagaagccaaagatctatctagaggaaggaagcagctgagcgggt 5141
 QY 4482 gccaaagcgagcgagccagcgctgttaacctctgtcctataatagaagccaaagccca 4541
 Db 5142 gccaaagcgagcgagccagcgctgttaacctctgtcctataatagaagccaaagccca 5201
 QY 4542 gaaccagaagcagatatacccttagaagactggaagataaaagaatagacagcttc 4601
 Db 5202 gaaccagaagcagatatacccttagaagactggaagataaaagaatagacagcttc 5261
 QY 4602 tctgaagactcggagggagcctgtctatccatatagtggaaggaatcctgcccacaaa 4661
 Db 5262 tctgaagactcggagggagcctgtctatccatatagtggaaggaatcctgcccacaaa 5321
 QY 4662 gaaggggttagaatalgttccaacagatatacgttctaaccaccttagaactaaacactg 4721
 Db 5322 gaaggggttagaatalgttccaacagatatacgttctaaccaccttagaactaaacactg 5381
 QY 4722 cagcaggttgtagaagacatcccttactatgttctggaagctcagagagtgagctagc 4781
 Db 5382 cagcaggttgtagaagacatcccttactatgttctggaagctcagagagtgagctagc 5441
 QY 4782 gtgtgtaaaacattgtgtccctcagcagctggttaatgtatcctccagaatattccca 4841
 Db 5442 gtgtgtaaaacattgtgtccctcagcagctggttaatgtatcctccagaatattccca 5501
 QY 4842 ggaagaagactaaagggaagcccaagcgcctcactggaagtgagacttactagta 4901
 Db 5502 ggaagaagactaaagggaagcccaagcgcctcactggaagtgagacttactagta 5561
 QY 4902 aagccggttaataatcggaaacaatatctatgttctgtagaaccttcttagaagtg 4961
 Db 5562 aagccggttaataatcggaaacaatatctatgttctgtagaaccttcttagaagtg 5621
 QY 4962 gtagaggttacttactaagaagaagacttcaacgctgtgtgcttaagaataactac 5021
 Db 5622 gtagaggttacttactaagaagaagacttcaacgctgtgtgcttaagaataactac 5681
 QY 5022 gaaattttccaagaattggaatacctaaggttaataggtcagaacaatgttccagcttc 5081
 Db 5682 gaaattttccaagaattggaatacctaaggttaataggtcagaacaatgttccagcttc 5741

QY 5082 gttgcccagtgatgcaaggagactggccaagatatgtgggattgtatgtgaacatgcatgt 5141
 Db 5742 gttgcccagtgatgcaaggagactggccaagatatgtgggattgtatgtgaacatgcatgt 5801
 QY 5142 gcatagaagcccaagagctcaggaagagtgtagagagatgataagacattaaagagacc 5201
 Db 5802 gcatagaagcccaagagctcaggaagagtgtagagagatgataagacattaaagagacc 5861
 QY 5202 cttac 5206
 Db 5862 cttac 5866

RESULT 11

AAC67032 standard; DNA; 6076 BP.
 ID AAC67032 standard; DNA; 6076 BP.
 XX
 AC AAC67032;
 XX
 PT 27-MAR-2001 (first entry)
 DE PERV env protein coding sequence spq ID NO: 32.
 DE
 KW Xenotransplantation; infectious agent; vaccine; ds.
 OS Porcine endogenous retrovirus.
 OS
 PN WO2000711726-A1.
 PN
 PD 30-NOV-2000.
 PD
 XX 24-MAY-2000; 2000WO-US14296.
 PF
 XX 24-MAY-1999; 99US-0135631.
 PR
 XX (MAYO-) MAYO MEDICAL VENTURES.
 PA
 XX Federspiel MJ;
 PI
 DR WPI; 2001-032041/04.
 DR
 XX
 XX Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PT
 PS Claim 16; Page 117-119; 144pp; English.
 PS
 XX
 CC The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 CC
 XX
 SQ Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;

Query Match

Best Local Similarity 98.7%; Pred No. 7.7e-278;
 Matches 2196; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 2992 tttcgtcgtgataatcacccacactagcaaacacattttgcttgcgtgaatgagagatcca 3041
 Db 3577 tttcgtcgtgataatcacccacactagcaaacacattttgcttgcgtgaatgagagatcca 3636
 QY 3042 ggtacgggaagaacccggcagcttcaacttgagaccgagctggcccaaggttcaagaactcc 3101
 Db 3637 ggtacgggaagaacccggcagcttcaacttgagaccgagctggcccaaggttcaagaactcc 3696
 QY 3102 ccgacacatcttgacgaagccctacacagaggaactgtgccaacttcagatccaacacct 3161
 Db 3697 ccgacacatcttgacgaagccctacacagaggaactgtgccaacttcagatccaacacct 3756

[illegible]

QY	4242	aaatccataacatttatacgcacgagtgatgcctcttcgcagtcgcacgtacacagg	4301
Db	4837	aaatccataacatttatacgcacgagtgatgcctcttcgcagtcgcacgtacacagg	4836
QY	4302	gcacatcataaacaagggtgtgtgtcttaccttaagcaggggaaggaataaagaacaagg	4361
Db	4897	gcacatcataaagaagggtgtgtgtcttaccttaagcaggggaaggaataaagaacaagg	4956
QY	4362	gaattctaaagcctattaaagccttaacatttgccaaaaagctagctattatacctgt	4421
Db	4957	gaattctaaagcctattaaagccttaacatttgccaaaaagctagctattatacctgt	5016
QY	4422	cctgcacatcaagaagcccaaaagatctcatcttaagaggaaacagatgctgcacagggtc	4481
Db	5017	cctgcacatcaagaagcccaaaagatctcatcttaagaggaaacagatgctgcacagggtc	5076
QY	4482	gccaaagcagcagcccgagctgtttaacctcttcgctcataalagaagaagcccaagcccca	4541
Db	5077	gccaaagcagcagcccgagctgtttaacctcttcgctcataalagaagaagcccaagcccca	5136
QY	4542	gaaccocagacgacagtaacacctgaagacacgycgaagagataaaaagaatagaccagttc	4601
Db	5137	gaaccocagacgacagtaacacctgaagacacgycgaagagataaaaagaatagaccagttc	5196
QY	4602	tcttgagctccggaggggagccttgctatacctcataatvggaaggaaatcctgtcccccacaa	4661
Db	5197	tcttgagctccggaggggagccttgctatacctcataatvggaaggaaatcctgtcccccacaa	5256
QY	4662	gaagggcttagaataatgtcccaacagatcatcgcgtctaacccacacttagaagaaataacactg	4721
Db	5257	gaagggcttagaataatgtcccaacagatcatcgcgtctaacccacacttagaagaaataacactg	5316
QY	4722	cagcagttgtgcagacaacatcccttctcaatgcttcgaagctacacagagtggtgcactcg	4781
Db	5317	cagcagttgtgcagacaacatcccttctcaatgcttcgaagctacacagagtggtgcactcg	5376
QY	4782	gtgtgtcaacaactgtgtgtgcctgcagcgtgtgttaatgtcaatcccttcagaataactcca	4841
Db	5377	gtgtgtcaacaactgtgtgtgcctgcagcgtgtgttaatgtcaatcccttcagaataactcca	5436
QY	4842	ggaaaagagactaagggaagcccccacagcgcctcaactgggaagtgtgactcttagagta	4901
Db	5437	ggaaaagagactaagggaagcccccacagcgcctcaactgggaagtgtgactcttagagta	5496
QY	4902	aagccggtctaataacggaaacaataatctatgtgttttgttagaacacttccagagttg	4961
Db	5497	aagccggtctaataacggaaacaataatctatgtgttttgttagaacacttccagagttg	5556
QY	4962	gttagaggtcatctcctaagaagaagacttaaacgcyggtgtgtgtcgaagaataactvgag	5021
Db	5557	gttagaggtcatctcctaagaagaagacttaaacgcyggtgtgtgtcgaagaataactvgag	5616
QY	5022	gaatttttccaagaatttggaaatacctaaggtaataaggttacaagcaatgtgtccagcttc	5081
Db	5617	gaatttttccaagaatttggaaatacctaaggtaataaggttacaagcaatgtgtccagcttc	5676
QY	5082	gttgcccaagtaagtcagagactgcccgaagatatttggtgttgatggaaactgtactgt	5141
Db	5677	gttgcccaagtaagtcagagactgcccgaagatatttggtgttgatggaaactgtactgt	5736
QY	5142	gcatcagaaccccaagctcagacaggttagagagagatgaataagcaatataaagagacc	5201
Db	5737	gcatcagaaccccaagctcagacaggttagagagagatgaataagcaatataaagagacc	5796
QY	5202	cttac 5206	
Db	5797	cttac 5801	
RESULT 12			
AAAT74884			
ID AAAT74884 standard; cDNA: 7892 BP.			

QY	2973	gagcgcctctctgcctcgtagatataaccccctagccaaacactcttttgccttgaaag	3032
Db	2970	gagcgcctctctgcctcgtagatataaccccctagccaaacactctttgccttgaaag	3029
QY	3033	agagatccaggtatccgggaagaaacccggagactctacccctgaaacccgactgccccaaaggttc	3092
Db	3030	agagatccaggtatccgggaagaaacccggagactctacccctgaaacccgactgccccaaaggttc	3089
QY	3093	aagaaactccccgaacacattcttgacgaagcctctacacagggagacttggccaaattccaggtac	3152
Db	3090	aagaaactccccgaacacattcttgacgaagcctctacacagggagacttggccaaattccaggtac	3149
QY	3153	caaacaccttaaggttgacccctctccacagatcgtgagatgacactctctcggcggaagaccac	3212
Db	3150	caaacaccttaaggttgacccctctccacagatcgtgagatgacactctctcggcggaagaccac	3209
QY	3213	aaacagagactgctctagaaaggtatcagaagagcactactgtcgtgaattgtctgactaagctac	3272
Db	3210	aaacagagactgctctagaaaggtatcagaagagcactactgtcgtgaattgtctgactaagctac	3269
QY	3273	agagactctgcttaaggaagggcccaagattcttcagaaagagaggttaacataacttgggggtacat	3332
Db	3270	agagactctgcttaaggaagggcccaagattcttcagaaagagaggttaacataacttgggggtacag	3329
QY	3333	ttgcggggcgccgacgacatctgctcgaacgagacacggaaagaaactgtatgccagataccg	3392
Db	3330	ttgcggggcgccgacgacatctgctcgaacgagacacggaaagaaactgtatgccagataccg	3389
QY	3393	ggcccaacacacagccaaacaaggtgagagaaggtttttgggggaacagcctggagttttgcagactg	3452
Db	3390	ggcccaacacacagccaaacaaggtgagagaaggtttttgggggaacagcctggagttttgcagactg	3449
QY	3453	tggatcccggggttctgagaccttaagacagcccaactctacccgtctaaccaagaagaaaggg	3512
Db	3450	tggatcccggggttctgagaccttaagacagcccaactctacccgtctaaccaagaagaaaggg	3509
QY	3513	ggattctctctgggctcctcgtgagcaacagaaagcatttgatgctataaaagagccctctg	3572
Db	3510	ggattctctctgggctcctcgtgagcaacagaaagcatttgatgctataaaagagccctctg	3569
QY	3573	agcgacactgctctcggccctccctcgaagataacaaacctttacaccttlatctgatgtgagag	3632
Db	3570	agcgacactgctcctggccctccctcgaagataacaaacctttacaccttlatctgatgtgagag	3629
QY	3633	cgtaagaggtatgcccggagaggttttaaacccaaaccttagagccatatgagagagacctgct	3692
Db	3630	cgtaagaggtatgcccggagaggttttaaacccaaaccttagagccatatgagagagagctgct	3689
QY	3693	ggcctacgttcaaaagaaacttgatccctctgtagccaggtgttggccgtatgcttgtaagact	3752
Db	3690	ggcctacgttcaaaagaaacttgatccctctgtagccaggtgttggccgtatgcttgtaagact	3749
QY	3753	atcgagcgtgtgacatactgctataagagacgcttgacaatltgaacttttggacagataata	3812
Db	3750	atcgagcgtgtggtgacatactgctataagagacgcttgacaatltgaacttttggacagataata	3809
QY	3813	actgtataatgcccccatgacttgtagaatacgtctcggagagcccccaagccagatgagtg	3872
Db	3810	actgtataatgcccccatgacttgtagaatacgtctcggagagcccccaagccagatgagtg	3869
QY	3873	accaaagcccgatgagcaacatlatcaagacctgtgcttctcacaagagaggttcaacttgcct	3922
Db	3870	accaaagcccgatgagcaacatlatcaagacctgtgcttctcacaagagaggttcaacttgcct	3920
QY	3933	ccaaacgacgctctcaaaccttgcacactcttctgctctgaagaagctgtataaaccaatgact	3992
Db	3930	ccaaacgacgctctctcaaaccttgcacactcttctgctctgaagaagctgtataaaccaatgact	3989

[illegible]

Primer	Sequence	Position
Db	ctgcatctgtgcatacagaccccaagctcagacagcgtatagagagatgaatagaacatt	5130
QY	aaagagacccctaac	5193
Db	aaagagacccctaac	5190

RESULT 13

1D AAF77727 standard; DNA; 8132 BP.

AC AAF77727;
VV

23-MAY-2001 (first entry)

XX nucleotide sequence of a retrovirus found in miniature swine.

KW SS. xenotransplantation; miniature swine;

OS	Unidentified
xx	

PN US6190861-B1
XX

PD 20-FEB-2001.
XX

13-DEC-1996; 9605-0766528.
XX

14 DEC 1993; 9505-05/2645.
XX

(S) GEN HOSPITAL CORP.

XX

Fishman JA;

DR WPI; 2001-256211/26.
P-PSDB: AAB73395 AAB73396 AAB73397

Assessing risk of endocrine-related

or miniature swine retroviral genome - probe sequences derived from swine

Claim 1; Fig 3; 127pp; English.

The present invention relates to a method for screening a cell or tissue for the presence or expression of a retrovirus (RV), comprising contacting a target nucleic acid from the cell or tissue with a second nucleic acid from the present invention (e.g. the present sequence or a fragment thereof). The method is useful for RV detection and to assess graft transplantation risk. Screening of animals allows the elimination of donors with active replication of known viruses. Inactive proviruses can be detected and inactivated, allowing identification and elimination of potential human pathogens derived from swine in a manner not possible in the outbred human organ donor population and is important to the development of human xenotransplantation.

sequence 8132 BP; 2248 A; 1977 C; 2037 G; 1870 T; 0 other;

Query Match

Best Local Similarity 98.7%; Pred. No. 1.1e-261;
Matches 2204; Conservative 0; Mismatches 30; Indels 0; Gaps 0.

Y 2973 gatgccttctctgcctgagattacacccaactagccaacacttttgccttcgaatq 3032

29/0 gatgccttctctgcctgagattacacccactagccaaccacttttgcctcgatgg 3029

ayagaccaglacgggaacccggtcaccctggaccgactgcccaaggttc 3092

3089

[illegible]

Db 3090 aagaactcccgacacattcttgaagagccctacacagagccctggccaaacttcagatc 3149
QY 3153 caacacccctcaagtgaacctctccagctacgttgaagtgacctgtctcggcggaagccac 3212
Db 3150 caacacccctcaagtgaacctctccagctacgttgaagtgacctgtctcggcggaagccac 3209
QY 3213 aaacagagctgtctgaaggtacgaagagcactactgctggatgtctgacctagctac 3272
Db 3210 aaacagagctgtctgaaggtacgaagagcactactgctggatgtctgacctagctac 3269
QY 3273 agaacctctgtctgaagagccagatcttcagagagagagtgtaacatacttggggtacagt 3332
Db 3270 agaacctctgtctgaagagccagatcttcagagagagagtgtaacatacttggggtacagt 3329
QY 3333 ttgcggggcgggcagcagatgtgtacgagagcagagaaagaaagctgtacagataccg 3392
Db 3330 ttgcggggcgggcagcagatgtgtacgagagcagagaaagaaagctgtacagataccg 3389
QY 3393 gcccaacacacagcccaacaaagtgagagagtttttgggagacagctggatcttgcagactg 3452
Db 3390 gcccaacacacagcccaacaaagtgagagagtttttgggagacagctggatcttgcagactg 3449
QY 3453 tggatcccggtgttgcagacttgcagagcccaactacccgttaacacaaagaaagag 3512
Db 3450 tggatcccggtgttgcagacttgcagagcccaactacccgttaacacaaagaaagag 3509
QY 3513 ggaattccctgtgctcctcagcagcccaagcagaaagcatttgatctcaaaaagccctgtg 3572
Db 3510 ggaattccctgtgctcctcagcagcccaagcagaaagcatttgatctcaaaaagccctgtg 3569
QY 3573 aggcacacgtctgtgctcctcagcagcccaagcagaaagcatttgatctcaaaaagccctgtg 3632
Db 3570 aggcacacgtctgtgctcctcagcagcccaagcagaaagcatttgatctcaaaaagccctgtg 3629
QY 3633 cgttaagagagtgagcccgagagagttttaacccaacccaaagacatagagagagccctgtg 3692
Db 3630 cgttaagagagtgagcccgagagagttttaacccaacccaaagacatagagagagccctgtg 3689
QY 3693 gcttaacctgtcaagaagagcttgcctgtgaagcagtggtgtgacctgtctgaagagct 3752
Db 3690 gcttaacctgtcaagaagagcttgcctgtgaagcagtggtgtgacctgtctgaagagct 3749
QY 3753 atcgcagctgtgtgcaatactgtgtcaagagcgttgacaatgagacttgggagcaataata 3812
Db 3750 atcgcagctgtgtgcaatactgtgtcaagagcgttgacaatgagacttgggagcaataata 3809
QY 3813 actgttaagagcccaatgagcagatgtgtgagcaatcgttgcgagcccccagacagatgtg 3872
Db 3810 actgttaagagcccaatgagcagatgtgtgagcaatcgttgcgagcccccagacagatgtg 3869
QY 3873 accaagcccgatgaccccaataaagcctgtgtctcaacagagaggttcaacttgcct 3932
Db 3870 accaagcccgatgaccccaataaagcctgtgtctcaacagagaggttcaacttgcct 3929
QY 3933 ccacacagccgtctcacaacccctgtcacttctgtcgtgaagagagctgtgaacaaatgtact 3992
Db 3930 ccacacagccgtctcacaacccctgtcacttctgtcgtgaagagagctgtgaacaaatgtact 3989
QY 3993 catgtatgcatacaatactgtatgtgaagagagctgtgtgagcagagcttcaacagata 4052
Db 3990 catgtatgcatacaatactgtatgtgaagagagctgtgtgagcagagcttcaacagata 4049
QY 4053 ccgctgactgtgagagagtgtaactgtgtcactgtacgtgaagagagctgtgtgtgaaggt 4112
Db 4050 ccgctgactgtgagagagtgtaactgtgtcactgtacgtgaagagagctgtgtgtgaaggt 4109
QY 4113 aagagagatgt 4172
Db 4110 aagagagatgt 4169
QY 4173 ccggaagagagctcagcagaaagagctgagctcagctcagcagagcttggcggtg 4232
Db 4170 ccggaagagagctcagcagaaagagctgagctcagctcagcagagcttggcggtg 4229

QY 4233 gccgaagggaaatccataaactttaacgagacaggtatgtccttggcgactgcagac 4292
Db 4230 gccgaagggaaatccataaactttaacgagacaggtatgtccttggcgactgcagac 4289
QY 4293 gtacacggggcactataaacaaggggtgtgttactctcagcgagggagaaataag 4352
Db 4290 gtacacggggcactataaacaaggggtgtgttactctcagcgagggagaaataag 4349
QY 4353 aaacaagaggaattctcaagctttagaagccttcaacttgcacaaagagcttagctatt 4412
Db 4350 aaacaagaggaattctcaagctttagaagccttcaacttgcacaaagagcttagctatt 4409
QY 4413 atacaacttccgtgacacacagaaagccaaagatctatctagagggagaaagagatgtgt 4472
Db 4410 atacaacttccgtgacacacagaaagccaaagatctatctagagggagaaagagatgtgt 4469
QY 4473 gaccgggtgtgccaagcagagccagctgttgaacttgcctataatagaagagcc 4532
Db 4470 gaccgggtgtgccaagcagagccagctgttgaacttgcctataatagaagagcc 4529
QY 4533 aaagcccgagaaacccagacagatagacacccctagaagagctgtgcaagagataaagata 4592
Db 4530 aaagcccgagaaacccagacagatagacacccctagaagagctgtgcaagagataaagata 4589
QY 4593 gaccagttctctgagaccccgagagggagacgttataactataatgggaagaaatcctg 4652
Db 4590 gaccagttctctgagaccccgagagggagacgttataactataatgggaagaaatcctg 4649
QY 4653 ccccaaaagaaaggttagaataatgtccaacagatatacgtcttaacccactagagact 4712
Db 4650 ccccaaaagaaaggttagaataatgtccaacagatatacgtcttaacccactagagact 4709
QY 4713 aaacacctgacagagtggtgcagaacatcccttactgttctgagagctacagaggt 4772
Db 4710 aaacacctgacagagtggtgcagaacatcccttactgttctgagagctacagaggt 4769
QY 4773 gctgaactcgtgtgtcaaacatctgtgtccctcagctgtgttaatgttaacttccaga 4832
Db 4770 gctgaactcgtgtgtcaaacatctgtgtccctcagctgtgttaatgttaacttccaga 4829
QY 4833 atacctcagagaaagagactaaagggagagccacccagcgctcaactggaagtgagctc 4892
Db 4830 atacctcagagaaagagactaaagggagagccacccagcgctcaactggaagtgagctc 4889
QY 4893 actgaggttaagcccggtcaaaatacgaacaataatctatgttctgttagacactt 4952
Db 4890 actgaggttaagcccggtcaaaatacgaacaataatctatgttctgttagacactt 4949
QY 4953 tcaagatgtgtgaagagcttactcactaagaagaagagcttcaacccgtgtgtgtgaataa 5012
Db 4950 tcaagatgtgtgaagagcttactcactaagaagaagagcttcaacccgtgtgtgtgaataa 5009
QY 5013 atactgtgaagaaatttttccaagatttgaatactactaaggttaataggttcaagaatgt 5072
Db 5010 atactgtgaagaaatttttccaagatttgaatactactaaggttaataggttcaagaatgt 5069
QY 5073 ccagcttctgtgtgccaaggttaagtcagagagctgtgcaagatattgtgtgtgtgtgtgt 5132
Db 5070 ccagcttctgtgtgccaaggttaagtcagagagctgtgcaagatattgtgtgtgtgtgtgt 5129
QY 5133 ctgcattgtgtacatacagagcccaagctcagagacaggttagagagagatgaataagacatt 5192
Db 5130 ctgcattgtgtacatacagagcccaagctcagagacaggttagagagagatgaataagacatt 5189
QY 5193 aaagagacccctac 5206
Db 5190 aaagagacccctac 5203

RESULT 14
AAC67022
ID AAC67022 standard; DNA: 4918 BP.


```

Db 3656 agttcagcacaagaagctgagctcatggtccctcaagcgaagcttgcggtgccaagg 3715
QY 4242 aaatccataaacatttatacaggaagcaggatgaccttgcgactgcacactacacggg 4301
Db 3716 aaatccataaacatttatacaggaagcaggatgaccttgcgactgcacactacacggg 3775
QY 4302 gccatctataaacaagggtgtgcttaacctcagcagggaaggaaataagaacaagg 4361
Db 3776 gccatctataaacaagggtgtgcttaacctcagcagggaaggaaataagaacaagg 3835
QY 4362 gaaattcctaagcctatagaagccttacatttgcacaaaaggctagctattataactgt 4421
Db 3836 gaaattcctaagcctatagaagccttacatttgcacaaaaggctagctattataactgt 3895
QY 4422 cctggacatcagaagaagccaagatc 4446
Db 3896 cctggacatcagaagaagccaagatc 3920

```

Search completed: February 24, 2002, 07:42:18
 Job time: 16843 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:05:29 ; Search time 624.99 Seconds
(without alignments)
24.691 Million cell updates/sec

Title: US-09-171-553B-14
Perfect score: 18
Sequence: 1 ccacagtcgtacacacacg 18
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Description	
Result No.	Score	Match	Length	ID			
C 1	18	100.0	633	22	AAC67007	PERV-MSL LTR SEQ I	
C 2	18	100.0	704	22	AAC67006	IamAI LTR SEQ ID N	
C 3	18	100.0	3482	20	AAV82749	Pig endogenous ret	
C 4	18	100.0	4918	22	AAC67021	PERV env protein c	
C 5	18	100.0	6076	22	AAC67021	PERV env protein c	
C 6	18	100.0	6076	22	AAC67032	PERV env protein c	
C 7	18	100.0	7333	22	AAF77726	Defective retrovir	
C 8	18	100.0	7362	22	AAC67019	PERV env protein c	
C 9	18	100.0	7393	18	AAT74883	Porcine retrovirus	
C 10	18	100.0	7873	22	AAC67023	PERV env protein c	
C 11	18	100.0	7892	18	AAT74884	Miniature swine re	

C 12	18	100.0	8060	18	AAT74811	Porcine retrovirus	
C 13	18	100.0	8060	22	AAF77725	Tsukuba-1 cDNA. P	
C 14	18	100.0	8132	22	AAF77727	Nucleotide sequenc	
C 15	18	100.0	8196	18	AAV09699	Porcine retrovirus	
C 16	18	100.0	8209	18	AAV09700	Porcine retrovirus	
C 17	17	94.4	20	18	AAT74869	Porcine retrovirus	
C 18	17	94.4	20	22	AAF77785	PCR primer #58. U	
C 19	17	94.4	28	20	AAV82760	PCR primer PL148 u	
C 20	15.4	85.6	6511	16	AAQ95493	Human Cdn-2 DNA.	
C 21	14.8	82.2	285	22	AAH0831	Human tumour assoc	
C 22	14.8	82.2	752	22	AAH04528	Human cDNA clone (
C 23	14.8	82.2	935	21	AAV09918	Fusarium venenatum	
C 24	14.8	82.2	4299	19	AAV03310	Pyrolobus fumarius	
C 25	14.4	80.0	3178	20	AAV99865	Human secreted pro	
C 26	14.4	80.0	15894	19	AAV18265	Human secreted pro	
C 27	14.4	80.0	15894	20	AAZ22902	Measles virus 1977	
C 28	14.4	80.0	32249	22	AAI63718	Nucleotide sequenc	
C 29	13.8	76.7	130	21	AAC14197	Human kidney relat	
C 30	13.8	76.7	131	21	AAC29646	Human secreted pro	
C 31	13.8	76.7	155	21	AAC13905	Human secreted pro	
C 32	13.8	76.7	178	21	AAC13629	Human secreted pro	
C 33	13.8	76.7	198	21	AAC23537	Human secreted pro	
C 34	13.8	76.7	300	20	AAZ13103	Human gene express	
C 35	13.8	76.7	300	20	AAV98595	Human cancer cell	
C 36	13.8	76.7	318	21	AAC28819	Human secreted pro	
C 37	13.8	76.7	344	21	AAC32109	Human secreted pro	
C 38	13.8	76.7	391	22	AAV66266	Novel human polynu	
C 39	13.8	76.7	417	21	AAC24025	Human secreted pro	
C 40	13.8	76.7	432	22	AAH13493	Human secreted pro	
C 41	13.8	76.7	530	21	AAZ49231	Human cDNA clone (
C 42	13.8	76.7	537	22	AAH10051	Human cDNA clone (
C 43	13.8	76.7	583	22	AAH10021	Human cDNA clone (
C 44	13.8	76.7	585	22	AAH10314	Human cDNA clone (
C 45	13.8	76.7	739	22	AAH08480	Human cDNA clone (

ALIGNMENTS

RESULT 1
AAC67007/c
ID AAC67007 standard; DNA; 633 BP.
XX AAC67007;
XX
XX 27-MAR-2001 (first entry)
XX
XX PERV-MSL LTR SEQ ID NO: 7.
XX
XX Xenotransplantation; infectious agent; vaccine; ds.
XX Porcine endogenous retrovirus.
XX
XX WO200071726-A1.
XX
XX PD 30-NOV-2000.
XX
XX PF 24-MAY-2000; 2000WO-US14296.
XX
XX PR 24-MAY-1999; 99US-0135631.
XX
XX (MAYO-) MAYO MEDICAL VENTURES.
XX
XX Federspiel MJ;
XX
XX WPI; 2001-032041/04.
XX
XX Inhibiting or preventing infectious agent transmission in mammalian
XX transplant recipients, by introducing recombinant DNA comprising DNA
XX encoding extracellular proteins of the agent into donor cells, such as
XX swine cells -
XX
XX Disclosure; Page 98; 144pp; English.

XX The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 633 BP; 170 A; 125 C; 147 G; 191 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 633;
Best Local Similarity 100.0%; Pred. No. 2.4; 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 ccacagtcgtacaccacg 18
|||||
DB 497 CCACAGTCGTACACCACG 480

RESULT 2
AAC67006/c
ID AAC67006 standard; DNA; 704 BP.
XX
AC AAC67006;
XX
DT 27-MAR-2001 (first entry)
XX
DE 1amA1 LTR SEQ ID NO: 6.
XX
KW Xenotransplantation; infectious agent; vaccine; ds.
XX
OS Unidentified.
XX
PN WO200071726-A1.
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14296.
XX
PR 24-MAY-1999; 99US-0135631.
XX
PA (MAYO-) MAYO MEDICAL VENTURES.
XX
PI Federspiel MJ;
XX
DR WPI; 2001-032041/04.
XX
PT Inhibiting or preventing infectious agent transmission in mammalian
PT transplant recipients, by introducing recombinant DNA comprising DNA
PT encoding extracellular proteins of the agent into donor cells, such as
PT swine cells -
XX
PS Disclosure; Page 97-98; 144pp; English.
XX
CC The present invention provides a method to prevent the transmission of
CC infectious agents during xenotransplantation. This involves introducing
CC to donor swine cells a recombinant DNA encoding a peptide fragment from
CC the infectious agent, and then introducing these cells into the
CC transplant recipient.
XX
SQ Sequence 704 BP; 187 A; 162 C; 158 G; 197 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 704;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
|||||
DB 568 CCACAGTCGTACACCACG 551

RESULT 3

AAV82749/c
ID AAV82749 standard; DNA; 3482 BP.
XX
AC AAV82749;
XX
DT 25-FEB-1999 (first entry)
XX
DE Pig endogenous retrovirus (PERV)-B envelope (env) gene region.
XX
KW Pig endogenous retrovirus; PERV-A; envelope protein; PERV-B; subtype;
KW probe; primer; detection; retrovirus; human tissue; xenotransplant;
KW primary porcine tissue; human cell line; porcine cell line; ss.
XX
OS Pig endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT 911..2884
FT /*tag- a
FT /product= envelope_protein
XX
PN WO9853104-A2.
XX
PD 26-NOV-1998.
XX
PF 18-MAY-1998; 98WO-GB01428.
XX
PR 16-MAY-1997; 97GB-0010154.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Stoye JP, Weiss RA;
XX
DR WPI; 1999-045324/04.
DR P-PSDB; AAW85452.
XX
PT Newly isolated nucleic acid probe capable of hybridising to either
PT the PERV-A or PERV-B env gene - useful in the detection of
PT retroviruses, and their subtypes, in a sample of porcine/human
PT tissue.
XX
PS Claim 3; Page 21-23; 36pp; English.
XX
CC The present sequence encodes a pig endogenous retrovirus (PERV)-B
CC envelope protein. PERV exists in two different subtypes, PERV-A and
CC PERV-B. The differences are reflected in sequence divergence in the
CC envelope genes. Probes and primers can be derived from the envelope
CC (env) genes of PERV-A and PERV-B. The probes and primers are used in
CC a method to detect retroviruses in a sample of porcine/human tissue,
CC particularly primary porcine tissue and human cell lines that have been
CC cultivated in the presence of a porcine cell line, or human tissue from
CC a patient with a xenotransplant. Subtype of PERV in a sample containing
CC one of the PERV env genes can also be determined.
XX
SQ Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;

Query Match 100.0%; Score 18; DB 20; Length 3482;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
|||||
DB 3439 CCACAGTCGTACACCACG 3422

RESULT 4
AAC67022/c
ID AAC67022 standard; DNA; 4918 BP.
XX
AC AAC67022;
XX
DT 27-MAR-2001 (first entry)
XX

DE PERV env protein coding sequence SEQ ID NO: 22.
 XX xenotransplantation; infectious agent; vaccine; ds.
 XX Porcine endogenous retrovirus.
 XX WO200071726-A1.
 XX 30-NOV-2000.
 PD 24-MAY-2000; 2000WO-US14296.
 PF 24-MAY-1999; 99US-0135631.
 PR (MAYO-) MAYO MEDICAL VENTURES.
 PA Federspiel MJ;
 PI WPI; 2001-032041/04.
 XX

DR WPI; 2001-032041/04.
 XX Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PS Claim 16; Page 109-111; 144pp; English.
 XX The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX

SO Sequence 4918 BP; 1382 A; 1173 C; 1183 G; 1180 T; 0 other;
 XX

Query Match 100.0%; Score 18; DB 22; Length 4918;
 Best Local Similarity 100.0%; Pred. NO. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtaccacg 18
 ||||||||||||||||
 DB 4781 CCACAGTCGTACACACG 4764

RESULT 5
 AAC67021/C
 ID AAC67021 standard; DNA; 6076 BP.
 XX
 AC AAC67021;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE PERV env protein coding sequence SEQ ID NO: 21.
 XX
 KW Xenotransplantation; infectious agent; vaccine; ds.
 XX Porcine endogenous retrovirus.
 XX OS
 XX WO200071726-A1.
 XX 30-NOV-2000.
 PD 24-MAY-2000; 2000WO-US14296.
 PF 24-MAY-1999; 99US-0135631.
 PR (MAYO-) MAYO MEDICAL VENTURES.
 PA Federspiel MJ;
 PI WPI; 2001-032041/04.
 XX

PT Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PS Claim 16; Page 107-109; 144pp; English.
 XX The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX

SO Sequence 6076 BP; 1613 A; 1512 C; 1609 G; 1342 T; 0 other;
 XX

Query Match 100.0%; Score 18; DB 22; Length 6076;
 Best Local Similarity 100.0%; Pred. NO. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtaccacg 18
 ||||||||||||||||
 DB 617 CCACAGTCGTACACACG 600

RESULT 6
 AAC67032/C
 ID AAC67032 standard; DNA; 6076 BP.
 XX
 AC AAC67032;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE PERV env protein coding sequence SEQ ID NO: 32.
 XX
 KW Xenotransplantation; infectious agent; vaccine; ds.
 XX Porcine endogenous retrovirus.
 XX OS
 XX WO200071726-A1.
 XX 30-NOV-2000.
 PD 24-MAY-2000; 2000WO-US14296.
 PF 24-MAY-1999; 99US-0135631.
 PR (MAYO-) MAYO MEDICAL VENTURES.
 PA Federspiel MJ;
 PI WPI; 2001-032041/04.
 XX
 DR WPI; 2001-032041/04.
 XX Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -
 PS Claim 16; Page 117-119; 144pp; English.
 XX The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.
 XX

SO Sequence 6076 BP; 1619 A; 1505 C; 1598 G; 1354 T; 0 other;
 XX

Query Match 100.0%; Score 18; DB 22; Length 6076;
 Best Local Similarity 100.0%; Pred. NO. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||
 DB 617 CCACAGTGTACACACG 600

RESULT 7
 AAF77726/C

ID AAF77726 standard; DNA: 7333 BP.

AC AAF77726;

DT 23-MAY-2001 (first entry)

DE Defective retroviral genome isolated from PK-15 cell line.

KW Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line;
 ss.

OS Unidentified.

PN US6190861-B1.

PD 20-FEB-2001.

PF 13-DEC-1996; 96US-0766528.

PR 14-DEC-1995; 95US-0572645.

PA (GEHO) GEN HOSPITAL CORP.

PI Fishman JA;

DR WPI: 2001-256211/26.

DR P-PSDB: AAB73282, AAB73283, AAB73284.

PT Assessing risk of endogenous retroviruses in clinical practice and in
 xenotransplantation, comprises using probe sequences derived from swine
 or miniature swine retroviral genome -

PS Claim 1; Fig 2; 127pp; English.

CC The present invention relates to a method for screening a cell or tissue
 CC for the presence or expression of a retrovirus (RV), comprising
 CC contacting a target nucleic acid from the cell or tissue with a second
 CC nucleic acid from the present invention (e.g. the present sequence or a
 CC fragment thereof). The method is useful for RV detection and to assess
 CC of donors with active replication of known viruses. Inactive proviruses
 CC can be detected and inactivated, allowing identification and elimination
 CC of potential human pathogens derived from swine in a manner not possible
 CC in the outbred human organ donor population and is important to the
 CC development of human xenotransplantation.

SO Sequence 7333 BP; 1984 A; 1788 C; 1885 G; 1676 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 7333;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||

DB 7279 CCACAGTGTACACACG 7262

RESULT 8
 AAC67019/C

ID AAC67019 standard; DNA: 7362 BP.

AC AAC67019;

DT 27-MAR-2001 (first entry)

DE PERV env protein coding sequence SEQ ID NO: 19.

KW Xenotransplantation; infectious agent; vaccine; ds.

OS Porcine endogenous retrovirus.

PN WO200071726-A1.

PD 30-NOV-2000.

PE 24-MAY-2000; 2000WO-US14296.

PR 24-MAY-1999; 99US-0135631.

PA (MAYO-) MAYO MEDICAL VENTURES.

PI Federspiel MJ;

DR WPI: 2001-032041/04.

PT Inhibiting or preventing infectious agent transmission in mammalian
 PT transplant recipients, by introducing recombinant DNA comprising DNA
 PT encoding extracellular proteins of the agent into donor cells, such as
 PT swine cells -

PS Claim 16; Page 101-104; 144pp; English.

CC The present invention provides a method to prevent the transmission of
 CC infectious agents during xenotransplantation. This involves introducing
 CC to donor swine cells a recombinant DNA encoding a peptide fragment from
 CC the infectious agent, and then introducing these cells into the
 CC transplant recipient.

SO Sequence 7362 BP; 1997 A; 1821 C; 1881 G; 1663 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 7362;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 XX ||||||||||||||||

DB 682 CCACAGTGTACACACG 665

RESULT 9
 AAT74883/C

ID AAT74883 standard; cDNA: 7393 BP.

AC AAT74883;

DT 09-FEB-1998 (first entry)

DE Porcine retrovirus cDNA (defective).

KW Retrovirus; porcine; GAG protein; POL protein; ENV protein;
 KW xenotransplantation; infectious; provirus; organ transplant; donor;
 KW activated virus; PCR; ss.

OS Porcine retrovirus.

PH Key Location/Qualifiers

FT CDS 598..2172

FT mat_peptide 598..2169

FT CDS 2320..4737

FT mat_peptide 2320..3522

FT /tag= d

FT /note= "putative GAG protein"

FT /note= "putative POL coding region (partial) as
 described in the specification"

```

FT mat_peptide /note="putative POL protein (partial)"
FT 3516..4328 /*tag= e
FT /note="putative POL protein (partial)"
FT mat_peptide 4332..4748
FT /*tag= f
FT CDS 4738..6725
FT /*tag= g
FT /note="putative ENV coding region (partial) as
FT described in the specification"
FT mat_peptide 4752..6722
FT /*tag= h
FT /note="ENV protein (partial)"
FT W09721836-A1.
FT 19-JUN-1997.
FT 13-DEC-1996; 96WO-US19680.
FT 14-DEC-1995; 95US-0572645.
FT (GEHO ) GEN HOSPITAL CORP.
FT Fishman JA;
FT WPI; 1997-332804/30.
FT P-PSDB; AAW32091-W32095.
FT New nucleic acid from porcine retroviruses - used for detecting
FT viruses in transplant or other tissue and for assessing risk of
FT transmitting infection to graft recipient
FT Claim 16; Fig 2; 128bp; English.
XX
XX This cDNA sequence represents a defective purified swine retrovirus
XX found in PK-15 cells containing the putative coding regions for viral
XX GAG, POL and ENV proteins. There are a few in frame stop codons and
XX apparent frame shifts in the given coding sequence which alter features
XX of the translation. This sequence and PCR fragments generated from the
XX sequence (see AAT74812-774882) could be used to screen organs for the
XX presence of porcine retroviruses prior to xenotransplantation.
XX Transplantation can increase the likelihood of retroviral activation if
XX intact and infectious proviruses are present. The porcine retroviral
XX sequence can be used to generate probes to determine the level (e.g.
XX copy number) of intact (i.e. potentially replicating) porcine provirus
XX sequences in a strain of xenograft transplantation donors. It can be
XX used to detect mutations, genetic lesions or viral recombinants and to
XX determine the histological localisation of activated retroviruses. Using
XX Polymerase Chain Reaction DNA Quantitation (PQ) on blood mononuclear
XX cells, infectivity titration and susceptibility testing can be
XX performed. Ultimately animal donors without intact porcine retroviral
XX sequences or a lower copy number of viral elements could be selected.
SQ
SQ Sequence 7393 BP; 2004 A; 1801 C; 1896 G; 1692 T; 0 other;

```

```

Query Match 100.0%; Score 18; DB 18; Length 7393;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ccacagtcgtacaccacg 18
DB 7339 CCACAGTCTGACACG 7322

```

```

RESULT 10
AAC67023/C
ID AAC67023 standard; DNA: 7873 BP.
XX AAC67023;
XX 27-MAR-2001 (first entry)

```

```

XX PERV env protein coding sequence SEQ ID NO: 23.
DE Xenotransplantation; infectious agent; vaccine; ds.
XX
XX Porcine endogenous retrovirus.
XX W0200071726-A1.
XX 30-NOV-2000.
XX 24-MAY-2000; 2000WO-US14296.
XX 24-MAY-1999; 99US-0135631.
XX (MAYO-) MAYO MEDICAL VENTURES.
XX Federspiel MJ;
XX WPI; 2001-032041/04.
XX Inhibiting or preventing infectious agent transmission in mammalian
XX transplant recipients, by introducing recombinant DNA comprising DNA
XX encoding extracellular proteins of the agent into donor cells, such as
XX swine cells -
XX Claim 16; Page 112-115; 144bp; English.
XX
XX The present invention provides a method to prevent the transmission of
XX infectious agents during xenotransplantation. This involves introducing
XX to donor swine cells a recombinant DNA encoding a peptide fragment from
XX the infectious agent, and then introducing these cells into the
XX transplant recipient.
XX
SQ
SQ Sequence 7873 BP; 2200 A; 1913 C; 1962 G; 1798 T; 0 other;

```

```

Query Match 100.0%; Score 18; DB 22; Length 7873;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ccacagtcgtacaccacg 18
DB 7708 CCACAGTCTGACACG 7691

```

```

RESULT 11
AAT74884/C
ID AAT74884 standard; cDNA: 7892 BP.
XX AAT74884;

```

```

XX 09-FEB-1998 (first entry)

```

```

XX Miniature swine retrovirus cDNA.

```

```

XX Retrovirus: porcine; GAG protein; POL protein; ENV protein;
XX xenotransplantation; infectious; provirus; organ transplant; donor;
XX activated virus; PCR; ss.

```

```

XX Porcine retrovirus.

```

```

XX Key Location/Qualifiers
XX CDS 585..2159

```

```

XX mat_peptide

```

```

XX /*tag= a
XX 585..2156
XX /*tag= b
XX /note="putative GAG protein"

```

```

XX CDS 2307..5744
XX /*tag= c
XX 2307..5741

```

```

XX mat_peptide
XX /*tag= d
XX /note="putative POL protein"

```

```

FT CDS 5620..7536
FT /*tag= e
FT mat_peptide 5620..7533
FT /*tag= f
FT /note= "putative ENV protein"
PN WO9721836-A1.
PD 19-JUN-1997.
PF 13-DEC-1996; 96WO-US19680.
PR 14-DEC-1995; 95US-0572645.
PA (GENO ) GEN HOSPITAL CORP.
PI Fishman JA;
DR WPI; 1997-332804/30.
DR P-PSDB; AAM32096-W32098.
XX
XX New nucleic acid from porcine retroviruses - used for detecting
PT viruses in transplant or other tissue and for assessing risk of
PT transmitting infection to graft recipient
PS Claim 22; Fig 3; 128pp; English.
XX
XX This cDNA sequence represents a porcine retrovirus from miniature swine
CC containing the putative coding regions for viral GAG, POL and ENV
CC proteins. This sequence and PCR fragments generated from it
CC (see AAT74812-T74882) can be used to screen organs for the presence of
CC porcine retroviruses prior to xenotransplantation. Transplantation can
CC increase the likelihood of retroviral activation if intact and
CC infectious proviruses are present. The porcine retroviral sequence can be
CC used to generate probes to determine the level (e.g. copy number) of
CC intact (i.e. potentially replicating) porcine provirus sequences in a
CC strain of xenograft transplantation donors. It can be used to detect
CC mutations, genetic lesions or viral recombinants and also to determine
CC the histological localization of activated retrovirus. Using Polymerase
CC Chain Reaction DNA Quantitation (PQ) on blood mononuclear cells,
CC infectivity titration and susceptibility testing can be performed.
CC Ultimately animal donors without intact porcine retroviral sequences or a
CC lower copy number of viral elements could be selected.
XX
XX Sequence 7892 BP; 2191 A; 1915 C; 1980 G; 1806 T; 0 other;
SQ
Query Match 100.0%; Score 18; DB 18; Length 7892;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccacagtcgtacaccag 18
DB 19 CCACAGTCGTACACCAG 2
RESULT 12
AAT74811/C
ID AAT74811 standard; cDNA; 8060 BP.
XX
XX AAT74811;
XX
XX 11-FEB-1998 (first entry)
XX
XX Porcine retrovirus Tsukuba-1 cDNA.
XX
XX Retrovirus; porcine; GAG protein; POL protein; ENV protein;
KW xenotransplantation; infectious; provirus; organ transplant; donor;
KW activated virus; Tsukuba-1; PCR; ss.
XX
XX Porcine retrovirus.
OS
XX
XX Key Location/Qualifiers

```

```

FT CDS 86..2002
FT /*tag= a
FT /note= "putative ENV protein"
FT CDS 3112..4686
FT /*tag= b
FT /note= "putative GAG protein"
FT CDS 4871..8060
FT /*tag= c
FT /note= "putative POL protein (partial)"
PN WO9721836-A1.
PD 19-JUN-1997.
PF 13-DEC-1996; 96WO-US19680.
PR 14-DEC-1995; 95US-0572645.
PA (GENO ) GEN HOSPITAL CORP.
PI Fishman JA;
DR WPI; 1997-332804/30.
XX
XX New nucleic acid from porcine retroviruses - used for detecting
PT viruses in transplant or other tissue and for assessing risk of
PT transmitting infection to graft recipient
PS Claim 1; Fig 1; 128pp; English.
XX
XX This sequence represents the purified porcine retroviral cDNA
CC sequence of Tsukuba-1 and contains the putative coding regions for viral
CC proteins GAG, POL and ENV. This sequence and PCR fragments generated
CC from the sequence (see AAT74812-T74882) could be used to screen organs
CC for porcine retroviruses prior to xenotransplantation. Transplantation
CC can increase the likelihood of retroviral activation if intact and
CC infectious proviruses are present. The porcine retroviral sequence can be
CC used to generate probes to determine the level (e.g. copy number) of
CC intact (i.e. potentially replicating) porcine provirus sequences in a
CC strain of xenograft transplantation donors. It can be used to detect
CC mutations, genetic lesions or viral recombinants and to determine the
CC histological localization of activated retrovirus. Using Polymerase Chain
CC Reaction DNA Quantitation (PQ) on blood mononuclear cells, infectivity
CC titration and susceptibility testing can be performed. Ultimately animal
CC donors without intact porcine retroviral sequences or with a lower copy
CC number of viral elements could be selected.
XX
XX Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;
SQ
Query Match 100.0%; Score 18; DB 18; Length 8060;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccacagtcgtacaccag 18
DB 2546 CCACAGTCGTACACCAG 2529
RESULT 13
AAF77725/C
ID AAF77725 standard; cDNA; 8060 BP.
XX
XX AAF77725;
XX
XX 23-MAY-2001 (first entry)
XX
XX Tsukuba-1 cDNA.
XX
XX Tsukuba-1; retrovirus; graft transplantation; xenotransplantation; ss.
KW Tsukuba-1; retrovirus; graft transplantation; xenotransplantation; ss.
XX
XX Porcine retrovirus.
OS
XX
XX

```


PN US6190861-B1.
 PD 20-FEB-2001.
 XX
 PF 13-DEC-1996; 96US-0766528.
 XX
 PR 14-DEC-1995; 95US-0572645.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Fishman JA;
 XX
 DR WPI: 2001-256211/26.
 XX
 PT Assessing risk of endogenous retroviruses in clinical practice and in
 PT xenotransplantation, comprises using probe sequences derived from swine
 XX or miniature swine retroviral genome -
 PS Claim 1; Fig 1; 127pp; English.
 CC
 CC The present invention relates to a method for screening a cell or tissue
 CC for the presence or expression of a retrovirus (RV), comprising
 CC contacting a target nucleic acid from the cell or tissue with a second
 CC nucleic acid from the present invention (e.g. the present sequence or a
 CC fragment thereof). The method is useful for RV detection and to assess
 CC of donors with active replication of known viruses. Inactive proviruses
 CC can be detected and inactivated, allowing identification and elimination
 CC of potential human pathogens derived from swine in a manner not possible
 CC in the outbred human organ donor population and is important to the
 CC development of human xenotransplantation.
 SQ Sequence 8060 BP; 2233 A; 1959 C; 2012 G; 1856 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 8060;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 ||||||||||||||||
 DB 2546 CCACAGTCGTACACCAG 2529

RESULT 14
 AAF77727/c
 ID AAF77727 standard; DNA; 8132 BP.
 XX
 AC AAF77727;
 XX
 DE 23-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a retrovirus found in miniature swine.
 KW
 KW Retrovirus; graft transplantation; xenotransplantation; miniature swine;
 KM ss.
 XX
 OS Unidentified.
 OS
 PN US6190861-B1.
 XX
 PD 20-FEB-2001.
 XX
 PF 13-DEC-1996; 96US-0766528.
 XX
 PR 14-DEC-1995; 95US-0572645.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Fishman JA;
 XX
 DR WPI: 2001-256211/26.
 DR P-PSDB; AAB73285; AAB73286; AAB73287.

XX
 PT Assessing risk of endogenous retroviruses in clinical practice and in
 PT xenotransplantation, comprises using probe sequences derived from swine
 PT or miniature swine retroviral genome -
 XX
 PS Claim 1; Fig 3; 127pp; English.
 CC
 CC The present invention relates to a method for screening a cell or tissue
 CC for the presence or expression of a retrovirus (RV), comprising
 CC contacting a target nucleic acid from the cell or tissue with a second
 CC nucleic acid from the present invention (e.g. the present sequence or a
 CC fragment thereof). The method is useful for RV detection and to assess
 CC of donors with active replication of known viruses. Inactive proviruses
 CC can be detected and inactivated, allowing identification and elimination
 CC of potential human pathogens derived from swine in a manner not possible
 CC in the outbred human organ donor population and is important to the
 CC development of human xenotransplantation.
 SQ Sequence 8132 BP; 2248 A; 1977 C; 2037 G; 1870 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 8132;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
 ||||||||||||||||
 DB 8080 CCACAGTCGTACACCAG 8063

RESULT 15
 AAV09699/c
 ID AAV09699 standard; DNA; 8196 BP.
 XX
 AC AAV09699;
 XX
 DE 19-MAY-1998 (first entry)
 XX
 DE Porcine retrovirus DNA encoding, GAG, POL and ENV.
 XX
 KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
 KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
 XX
 OS Porcine retrovirus.
 XX
 FH
 FH Key Location/Qualifiers
 FT CDS 576..2126
 FT /*tag= a
 FT /product= GAG protein
 FT /note="viral core polypeptide"
 FT 2143..5733
 FT /*tag= b
 FT /product= POL protein
 FT /note="polymerase peptide sequence as given in
 FT specification"
 FT 5606..7576
 FT /*tag= c
 FT /product= ENV protein
 FT /note="envelope protein"
 XX
 MO9740167-A1.
 XX
 PD 30-OCT-1997.
 XX
 PF 18-APR-1997; 97WO-GB01087.
 XX
 PR 10-FEB-1997; 97GB-0002668.
 XX
 PR 19-APR-1996; 96GB-0008164.
 XX
 PA (IMUT-) IMUTRAN LTD.
 PA (ONE-) Q-ONE BIOTECH LTD.
 XX

PI Galbraith DN, Haworth C, Lees GM, Smith KT;
 XX WPI, 1997-535851/49.
 DR
 XX Polynucleotide encoding porcine retrovirus expression product -
 PT useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 XX
 PS Claim 4; Fig 2; 69pp; English.
 XX
 CC This DNA sequence encodes the porcine retrovirus (PoEV) virion core
 CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
 CC proteins can be used to develop viral vaccines, antisense nucleic acids,
 CC ribozymes and other antiviral agents. They can also be used in
 CC xeno-transplantation technology and as diagnostic tools.
 XX
 SQ Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other;

Query Match 100.0%; Score 18; DB 18; Length 8196;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccacagtcgtacaccag 18
 ||||||||||||||||
 DB 8130 CCACAGTCGTACACACG 8113

Search completed: February 24, 2002, 03:05:32
 Job time: 19498 sec

Mon Feb 25 07:44:48 2002

us-09-171-553b-1.feb2201i.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:56:31 ; Search time 197.25 Seconds
(without alignments)
3811.949 Million cell updates/sec

Title: US-09-171-553b-1

Perfect score: 3320
Sequence: 1 gaattcgagcgctgcac.....aaaaaaaaaaaaaaaa 3320

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 11338999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1091	32.9	7333	4 US-08-766-528-2	Sequence 2, Appli
2	786	23.7	2462	3 US-09-111-085-1	Sequence 1, Appli
3	704	21.2	8132	4 US-08-766-528-3	Sequence 3, Appli
4	375	11.3	8060	4 US-08-766-528-1	Sequence 1, Appli
5	321	9.7	3482	3 US-09-111-085-3	Sequence 3, Appli
6	265	8.0	445	4 US-09-376-781-7	Sequence 7, Appli
7	35	1.1	8655	3 US-09-352-990-7	Sequence 1, Appli
8	30	0.9	1279	4 US-09-277-716-31	Sequence 31, Appli
9	30	0.9	735	1 US-08-798-000-3	Sequence 3, Appli
10	29	0.9	1198	3 US-09-248-335-27	Sequence 27, Appli
11	29	0.9	1234	1 US-08-798-000-1	Sequence 1, Appli
12	29	0.9	2338	4 US-09-232-200-66	Sequence 66, Appli
13	29	0.9	2338	4 US-09-232-197-66	Sequence 66, Appli
14	29	0.9	7308	4 US-09-011-745-3	Sequence 4, Appli
15	29	0.9	7308	4 US-09-011-745-4	Sequence 4, Appli
16	29	0.9	7616	4 US-09-011-745-2	Sequence 13, Appli
17	29	0.9	8202	1 US-08-258-420-13	Sequence 1, Appli
18	29	0.9	8332	1 US-08-850-961-1	Sequence 5, Appli
19	29	0.9	140	1 US-08-628-417-5	Sequence 34, Appli
20	28	0.8	216	1 US-08-668-878A-34	Sequence 6, Appli
21	28	0.8	240	1 US-08-628-417-6	Sequence 24, Appli
22	28	0.8	330	4 US-09-078-294-24	Sequence 278, Appli
23	28	0.8	595	4 US-09-385-982-278	Sequence 18, Appli
24	28	0.8	698	4 US-09-376-781-18	Sequence 1, Appli
25	28	0.8	719	2 US-08-375-346A-1	Sequence 1, Appli
26	28	0.8	719	2 US-08-467-123B-1	Sequence 1, Appli
27	28	0.8	719	2 US-08-467-123B-1	Sequence 1, Appli

28	0.8	949	5 PCT-US91-08254-1	Sequence 1, Appli
29	0.8	949	5 PCT-US91-08254-2	Sequence 2, Appli
30	0.8	1132	3 US-08-894-731-3	Sequence 3, Appli
31	0.8	1578	4 US-09-416-050A-1	Sequence 1, Appli
32	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
33	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
34	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
35	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
36	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
37	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
38	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
39	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
40	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
41	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
42	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
43	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
44	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli
45	0.8	1578	4 US-09-664-800-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-2
Sequence 2, Application US/08766528

GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766, 528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572, 645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7333 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-2

Query Match 32.9%; Score 1091; DB 4; Length 7333;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1291; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1364 aaagggggttgcctccacagaggaataaagaaaggaattcctaagc 1423
|||||

QY	2504	accacgaggagactctgcaagttccacatcgcttccaaagtgtgagaattcagctatgta	256
Db	4657	ACTCAGGAGGAGACCTTCCAAgTTCACATCGCTTCCAAgTGGAGATTAGCTATGTTA	4718
QY	2564	gacgcacccgtgcaggaaccttgagactcgtgtgtgagaagaccattatctcgactcttga	2623
Db	4717	GACCCACCGTGCAGGAACCTCGAAGCTCGTAGAAGGACCTATCTCGACTTTGA	4776
QY	2624	ccacaccaacgctgttaagaatcgaaagatcccc	2658
Db	4777	CCACACCAACGGCTGTGAAGATGCAAGGATCCCC	4811

RESULT 2
 US-09-111-085-1
 ; Sequence 1, Application US/09111085
 ; Patent No. 6100034
 ; GENERAL INFORMATION:
 ; APPLICANT: Stoye, Jonathan P
 ; APPLICANT: Weiss, Robin A
 ; TITLE OF INVENTION: Detection of retroviral
 ; FILE REFERENCE: 4238/75168
 ; CURRENT APPLICATION NUMBER: US/09/111,085
 ; CURRENT FILING DATE: 1998-07-07
 ; EARLIER APPLICATION NUMBER: GB 9710154.7
 ; EARLIER FILING DATE: 1997-05-16
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2462
 ; TYPE: DNA
 ; ORGANISM: Porcine retrovirus
 ; US-09-111-085-1

Query Match	Local Similarity	23.7%	Score 786;	DB 3;	Length 2462;
Best Match	Similarity	99.98%	Pred. No. 0;		
Matches	836;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
QY 2459	tcgagtcggg tgaagccagcagctgtggaacacatcccggaagcctcctcagaagagact				
Db 1	tcgagtcggg tgaagccagcagctgtggaagcactccgggagcctcctcagaagagagact				
QY 2519	tgcaggtccacacatcgctctccaaagtctggaatctcagtcatactgtagaccacccgttcag				
Db 61	tgcaggtccacacatcgctctccaaagtctggaatctcagtcatactgtagagccacccgttcag				
QY 2579	gaaacctcgagacatcgctggggaagggacctatctccgtactcttgcacacacacacagctg				
Db 121	gaaacctcgagacatcgctggggaagggacctatctccgtactcttgcacacacacacagctg				
QY 2639	tgaagtcgaaagagatcccccaccctggatccatctgatatcccgctaaacggcgccacctc				
Db 181	tgaagtcgaaagagatcccccaccctggatccatctgatatcccgctaaacggcgccacctc				
QY 2699	ccgattcggggttggaagccgaaagactgaanaatcccccctaaagcttcgactcatctgcg				
Db 241	ccgattcggggttggaagccgaaagactgaanaatcccccctaaagcttcgactcatctgcg				
QY 2759	tgcgtccttactctgttaataaactccctccaagttaaatgtaaaacgctctgtgaacagcccg				
Db 301	tgcgtccttactctgttaataaactccctccaagttaaatgtaaaacgctctgtgaacagcccg				
QY 2819	aactccataaaacccatctatctccactcgtttacttaccagctccggtacaggtataat				
Db 361	aactccataaaacccatctatctccactcgtttacttaccagctccggtacaggtataat				
QY 2879	atataaagcactaaagggaagctccctctggggagcttggccttgaattatattctgc				
Db 421	atataaagcactaaagggaagctccctctggggagcttggccttgaattatattctgc				

QY 2939 ctctgacagtaacccctgtgtctcaatgacagagccacccccgagatgtaactccgtgct 2998
Db 481 ctctgacagtaacccctgtgtctcaatgacagagccacccccgagatgtaactccgtgct 540
QY 2999 taagggtttacgtttgtcccaagagaccccaataatgaagaatattgtgnaaatctccag 3058
Db 541 taagggtttacgtttgtcccaagagaccccaataatgaagaatattgtgnaaatctccag 600
QY 3059 gatttccttgcgaagaatgagagctgataactcttaataatgaggaattgtgnaatgcca 3118
Db 601 gatttccttgcgaagaatgagagctgataactcttaataatgaggaattgtgnaatgcca 660
QY 3119 gtctctgaagcaagagatgaatgaatgaatgttgaacaaatccctaccagttatacaca 3178
Db 661 gtctctgaagcaagagatgaatgaatgaatgttgaacaaatccctaccagttatacaca 720
QY 3179 tttaattatgcatgagagatgagaaagattgtgcaacagcggtgtacaaaagaatgtacga 3238
Db 721 tttaattatgcatgagagatgagaaagattgtgcaacagcggtgtacaaaagaatgtacga 780
QY 3239 aataagcaataagctgtcattgtgtgacactagattacttaaaataagtttcaact 3295
Db 781 aataagcaataagctgtcattgtgtgacactagattacttaaaataagtttcaact 837

RESULT 3

US-08-766-528-3
Sequence 3, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SPOUNCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-3

Query Match 21.2%; Score 704; DB 4; Length 8132;
Best Local Similarity 98.7%; Pred. No. 2.8e-299;
Matches 2204; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 22 gatgccttctctgcctgagatlaacacccactagaccaacacttttgcccttgcaatgg 81
Db 2970 gatgccttctctgcctgagatlaacacccactagaccaacactttttgcttccgaatgg 3029
QY 82 agagatccaggtacgaggaacccgagacgtctcaactgtgacccgactgtcccaaggttc 141
Db 3030 agagatccaggtacgaggaacccgagacgtctcaactgtgacccgactgtcccaaggttc 3089
QY 142 aagaactcccccagccactcttgaagagccctacacagagagacctgtgccaacttcagatc 201
Db 3090 aagaactcccccagccactcttgaagagccctacacagagagacctgtgccaacttcagatc 3149
QY 202 caacacccctcaagttgacccctccacagtaagttgagtaacctgtcttcgcygagaccac 261
Db 3150 caacacccctcaagttgacccctccacagtaagttgagtaacctgtcttcgcygagaccac 3209
QY 262 aaacaggaactgcttagaaggtacgaagacactctctggaattgtctgaactaggtac 321
Db 3210 aaacaggaactgcttagaaggtacgaagacactctctggaattgtctgaactaggtac 3269
QY 322 agagcctctgtctgtaagaagccagatttgcagagagagtaacataacttgggtacagt 381
Db 3270 agagcctctgtctgtaagaagccagatttgcagagagagtaacataacttgggtacagt 3329
QY 382 ttgcgggycgagcagatgtgtgacgagcagcaggaagaaactgttagtccagataccg 441
Db 3330 ttgcgggycgagcagatgtgtgacgagcagcaggaagaaactgttagtccagataccg 3389
QY 442 gcccccaacacagaccacaaagtgagaggttlttgggagcagctgattttgcaacagc 501
Db 3390 gcccccaacacagaccacaaagtgagaggttlttgggagcagctgattttgcaacagc 3449
QY 502 tggatcccgagggtgtgcagacttagcagcccccactacccgttaaccaaagaagaggg 561
Db 3450 tggatcccgagggtgtgcagacttagcagcccccactacccgttaaccaaagaagaggg 3509
QY 562 ggaattcctctggtctcctgagcaacagaaagcaattgtatgatacaaaaagccctgtg 621
Db 3510 ggaattcctctggtctcctgagcaacagaaagcaattgtatgatacaaaaagccctgtg 3569
QY 622 aggcgaactgtctgtgcccctcctgagtaactaaaccccttaacttctgtgagtgag 681
Db 3570 aggcgaactgtctgtgcccctcctgagtaactaaaccccttaacttctgtgagtgag 3629
QY 682 cgttaaggaatgagccgagaggttlaaccccaacccctagagacatgaggaagagacctgt 741
Db 3630 cgttaaggaatgagccgagaggttlaaccccaacccctagagacatgaggaagagacctgt 3689
QY 742 gcttaccctgtcaagaagactgtatcctgttagcagagtggttgcgagctatgtcgaaggtc 801
Db 3690 gcttaccctgtgtcaagaagactgtatcctgttagcagagtggttgcgagctatgtcgaaggtc 3749
QY 802 atgcgaactgtgtgcatactgtgtcaagaagcgtgacaaatgtgacttggagaggaatata 861
Db 3750 atgcgaactgtgtgcatactgtgtcaagaagcgtgacaaatgtgacttggagaggaatata 3809
QY 862 actgttaatgccccccactgtatgtgagagacactgttcgagagcccccaacagactgtgatg 921
Db 3810 actgttaatgccccccactgtatgtgagagacactgttcgagagcccccaacagactgtgatg 3869
QY 922 accaagcccgcatgaccccaactacaaagctgttctacagagagaggtgcaacttgcgt 981
Db 3870 accaagcccgcatgaccccaactacaaagctgttctacagagagaggtgcaacttgcgt 3929
QY 982 ccaccagcagcttcaaacacctgtgacactctctgcttgaagagactgtatgaacagagttact 1041
Db 3930 ccaccagcagcttcaaacacctgtgacactctctgcttgaagagactgtatgaacagagttact 3989
QY 1042 catgatgtgcatcaactatgtatgtagagagactgtgtccgcaagagaccttaacagacata 1101
Db 3990 catgatgtgcatcaactatgtatgtagagagactgtgtccgcaagagacctttacagacata 4049
QY 1102 ccgctgactgtgagaagtgactaactgtgttcaactgtacggaagcagctatgtgtgtgaggt 1161

```

Db 4050 CCGCTGACTGAGAAAGTGTAACTGTTCACTGACGGAAGAGCTATGTGTGAAAGT 4109
Qy 1162 aagagatgctgctgggagcgagctgggagggagccgacatctgggagcgagctg 1221
Db 4110 AAGAGATGCTGGGGGGGGGGTGTGTGACGGGACCCGACGATGTGGGCCAGCGCTG 4169
Qy 1222 ccggaaggaactcaagcgcaaaagctgagctcatgagccctcaagagcttggcgctg 1281
Db 4170 CCGGAAGGAATTTTCAAGCAAAAGGCTAGCTCATGCGCCCTCAGCAAGCTTTGGGGCTG 4229
Qy 1282 gccgaaggaagaaatccataacattatagcagacagctgagcttggcgagcgac 1341
Db 4230 GCCGAAGGGAATTCATTAACATTTATACGACAGCAGGTATGCTTTGCGACTGCAAC 4289
Qy 1342 gtacaggggccaatcataaacaagggggtgttcttactcagcagaggaagaaataag 1401
Db 4290 GTACATGGGGCCATCTATTAACAAAGGGGGTGTGCTTACTCAGCAGGAGGGAATTAAG 4349
Qy 1402 aacaaagaggaatctcaagctatagcctatagaccttaccattggcaaaaagctagctatt 1461
Db 4350 AACAAAGAGAAATTTCTAAGCTTATTAAGACCGCTTACATTTTCCAAAAGGCTAGCTATT 4409
Qy 1462 atacactgtcctgacatcacaagaagccaaagatcctatcagagggaaacagctgct 1521
Db 4410 ATACACTGTCTTGACATCAAGCAAAAGCTAATCTCATATTCAGAGAAACCAAGTGGCT 4469
Qy 1522 gaccgggttgcgaagcagcagccagcgtgttaaccttgcctataatagaacgcc 1581
Db 4470 GACGGGGTTCAGAGCAGGCGACCCAGGGTGTAACTTTCCTTATTAATGAATGCC 4529
Qy 1582 aaagcccccagagcccaagcagcagtaaaccttagaagactggcagaagataaaagata 1641
Db 4530 AAAGCCCCAGAACCCAGACGACATTAACCTTAGAAGACGCGCAAGAGATAAAAGATA 4589
Qy 1642 gaccagttctctgagactccgaagggagcctgtataccctcatatgggaagaaatcgt 1701
Db 4590 GACCAGTTCTCTGAGACTCCGGAAGGAGACTGCTATACCTCAGATGGGAAAGAAATCTG 4649
Qy 1702 ccccaaaaagaggttagaataatgltcaacagatcatgltcaacccactagaagact 1761
Db 4650 CCCCACAAAGAGGGTTAGATATGTCCACAGATACATGCTTAACCCACCTAGAGACT 4709
Qy 1762 aaacactgcaagatgtgcagaaacatcccttatcatgttctgaggtaccagaagtg 1821
Db 4710 AAACACTGTGACGAGTGTGCAAAACATCCCTTATCATTTCTGAGGCTACCGAGAGTG 4769
Qy 1822 gctgactcgggtgtcaacaattgtgtccctgacagctggttaacttaactcctcaga 1881
Db 4770 GCTGACTCGGTGTCAACATTTGTGTGCTGCTGACAGTGTATATGCTTCCAGA 4829
Qy 1882 atacctcagaagaaagactaaagggaagccacccagcgctcactgggaagtgcattc 1941
Db 4830 ATCCCTCCAGGAGAGACTAAAGGGAAGCCACCGCGCTCAGTGGGAAGTGGACTTC 4889
Qy 1942 actgaagtaagccgctaaatcagaaacaatatctatgttctgttagaacctt 2001
Db 4890 ACTGAGGTAAAGCCGGCTAAATTCGGAACCAATACCTATGTGTTTGTAGACACTTT 4949
Qy 2002 tcaggaatgggtagagctatcctactaagaagagactcaacgttggcttagaaga 2061
Db 4950 TCAGGATGGGTAGAGCTTATCTACTAAGAAAGAGACTTCAACCTGGTGGCTAAAGAA 5009
Qy 2062 atactggaagaaatttccaaagatttgaatacctaaglaaaggtcagacaatgct 2121
Db 5010 ATACTGGAAGAAATTTTCCAAAGATTGGAATTAACCTAAGGTATAGGTCACACATGCT 5069
Qy 2122 caagcttctgtcccaagtaagtcagagactgcccagaatacttggagatgttgga 2181
Db 5070 CCAAGCTTTTGTGGCCAGGTAAAGTCAAGGACGCGCAAGATATTTGGGGATTGATGAAA 5129
Qy 2182 ctgcatgtgcatagaccaccaagctcagacaggttagaagatbaatagaacctt 2241

```

```

Db 5130 CTGCATTGTGATACAGAACCCCAAGCTCAGACAGGTAGAGAGATGATAGAACATT 5189
Qy 2242 aaagagacccttac 2255
Db 5190 AAAGAGACCTTAC 5203

```

RESULT 4

```

US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-1

```

Query Match 11.3%; Score 375; DB 4; Length 8060;

Best Local Similarity 96.7%; Pred. No. 3.5e-15; Matches 1395; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```

Qy 22 gatgcctctctgctctagattcaaccacactagccaacacttttgcttgaatgg 81
Db 5495 GATGCTCTCTCTGCTCTAGATTACACCCCACTAGCCACACTTTTGGCTTGAATGG 5554
Qy 82 agagatcaggtacgggaagaacccgggagctcaactggaacccgactgcccagaagttc 141
Db 5555 AGAGATCCAGGTACGGGAAGAACCGGAGCTCACCTGAGACCCGACAGTCCCAAGGCTTC 5614
Qy 142 aagaactcccgaccatcttggagaagcccttaacaaggagactgcccacttaagatc 201
Db 5615 AAGAACTCCCGACCACTTTTGAAGCAAGCCCTTACACAGACACCTGGCCAACCTCAGATC 5674
Qy 202 caaacacctgaagtgacctctcctcagtaagtgagtaagctgtcttgagggaagcacc 261
Db 5675 CAACACCTCGAGTACCCCTCTCTCAGTACGTGATATCTGCTTGGGGGGGACCAACC 5734
Qy 262 aaacagactgtctagaaggtacgaaggaactactgtctggaatgtctgacttagctac 321

```

Db 5735 AAGAGGACTGCTTAGAAGGACGACGACACTACTGCTGGAATTCGACACTAGGCTAC 5794
QY 322 agagcctctgtaagaagccagagattgcaagagagagtaacatactggggtaagc 381
Db 5795 AGAGCCTGCTGTAAGAAAGGCCAGATTGCAAGAGAGGTAACATCTGGGGTACAGT 5854
QY 382 ttgaggagcgagcagcagatgctacgagagcagcagaaacatgagtcagataccg 441
Db 5855 TTACGGGAGCGGAGCATGGCTCAGCAGGACGAGAAACTGTATGTCAGATACCG 5914
QY 442 gcccaaacacagccaaacaaagttagagagttttgggagcagctgtagatltgcaagctg 501
Db 5915 GCCCAACACAGCCAAACAAATGAGAGATTTTGGGAGAGCTGATTTTGGAGACTG 5974
QY 502 tggatcccgagggtttgcaaccttagcagcccccactaccgcgttaacaaagaaaggg 561
Db 5975 TGGATCCCGGGGTTGGACCTTAGCAGCCCACTACTACCGGCTAACCAAGAAAGAGG 6034
QY 562 ggatctcctctgagcctcagcagcagaaagcalttgatgtalcaaaaagccctctg 621
Db 6035 GAATTCCTCTGGCTCTGAGACACAGCAAGCATTTGATGCTATCAAAAAGGCCCTGCTG 6094
QY 622 agcgacacctctctgcccctccctgacgtaactaaaccttacccttattgtgtagtag 681
Db 6095 AGCGCACCCTGCTCTGGCCCTCCCTGACGTAACCTTAACCTTTATGATGATGAG 6154
QY 682 cgttaaggagtagcccgagagattttaaccaccccttagagccatgtagagagacctgt 741
Db 6155 CGTAAGGAGTAGCCCGGGGAGTTTAAACCAACCCCTAGACCATGAGAAAGACCTGTC 6214
QY 742 ggcctacctgtaaaagagctgtaacctgtagcagtagtctggcccagctgtcgaagct 801
Db 6215 GCGTACCTGTCAAGAAAGCTGATCGTGTAGCCAGTGTGGCCCATATGCTGSAAGCT 6274
QY 802 atcgcaagctgtgcccatactggttcaagaagcgtgcaaatgactttggagc-agaat 860
Db 6275 ATCCAGCTGTGGCATTACTGTGTCAGAGAGCTGTCAAAATGATTTGGGACAAAGATAT 6334
QY 861 aacgtataagcccccatgcatgtagagaaatcggttcgagcagcccccagacagatgat 920
Db 6335 AACTGTATAGCCCCCATGATGATGAGAACATGTTGCGCAGGCCCAACCCAGATGAT 6394
QY 921 gaccaagcccgcaatgacccactlcaaaagcgtctctcaagagagaggtlcaacttcgc 980
Db 6395 GACCAAGCCCGCATGACCCCATATCAAAAGCCTCTCTCAACAGAGGGTCAAGCTTGGC 6454
QY 981 tcaacagcagcgtctcaacctgcaactctctgctggaagagagctgtaaacagtgac 1040
Db 6455 TCCACCAACCGCTCTCAACCTGTCCACTTCTGCTGAAGAGACTGATGATGACAGTAC 6514
QY 1041 tcatgattgcatcaactatgattttagagagactgggttcgcaagagacattacagat 1100
Db 6515 TCATGATTGCCATCACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACAT 6574
QY 1101 accgctgactggaagatgtaacctggtltaactgacggaagcagcattgltgtggaag 1160
Db 6575 ACCGTGACTGAGAAAGTGTAACTGCTGTTCACTGACGGAAGACGATATGTTGGTGAAG 6634
QY 1161 taagaagatggtctgggagcagctggtgagcgggaacccgcaagctctgggacagagct 1220
Db 6635 TAAAGAGATGGCTGGGGGGGGGTGTGAGGGGAGCCGACGATCTGGGGCCACACCTT 6694
QY 1221 gccggaaggaactcagcagcaaaagcttagctcagctcagcagaaactttggagct 1280
Db 6695 GCCGGAGGAGACTTCAGCACAAAGGCTGAGCTCATGGCCCTCAAGGAGCTTTGGGGCT 6754
QY 1281 gccggaaggaatcacaataacatltlacgacagcaggtatgacctttgcagtgcaca 1340
Db 6755 GGCGGAAGGAATTCATTAACATTTTATACGACAGCAGGATATGCTTTGGAGACTGCACA 6814
QY 1341 cgtacaaggggcatcatataacaagaagggtgtgcttaactcagcaggaagggaataaa 1400
Db 6815 CGTACATGGGGCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGGGAATATAA 6874

QY 1401 gaacaaaggaatcttaagcctatagaagcc 1434
Db 6875 GAACAAGAGAGAAATTCTTAAGCCTATTAGAACCC 6908
RESULT 5
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Sloye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match 9.7%; Score 321; DB 3; Length 3482;
Best Local Similarity 99.2%; Pred. No. 1.7e-133;
Matches 521; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1767 cctgcagcagttggttagaacaatcccttatatgtctggaagctaccaggggtgctga 1826
Db 6 cctgcagcagttggttagaacaatcccttatatgtctggaagctaccaggggtgctga 65
QY 1827 ctcggtgctcaaacatctgtgcccgcagctgtgtaagtcaatccttcagaatacc 1886
Db 66 ctggtgtgctcaaacatctgtgcccgcagctgtgtaagtcaatccttcagaatacc 125
QY 1887 tccagaaagagactaagggaagccaccagcgctcaactgaggagtgactcaatga 1946
Db 126 tccagaaagagactaagggaagccaccagcgctcaactgaggagtgactcaatga 185
QY 1947 ggtaaagccggctaaataaggaacaatatcatgtgtttttagaacctttccag 2006
Db 186 ggtaaagccggctaaataaggaacaatatcatgtgtttttagaacctttccag 245
QY 2007 atggtagaggtctactactaagaagaagactcaaccgtgtgtgctaaagaaatact 2066
Db 246 atggtagaggtctactactaagaagaagactcaaccgtgtgtgctaaagaaatact 305
QY 2127 ttctgttgcacagtagtcaaggtggtccaaagatattggggattgatttgaacttga 2186
Db 366 ttctgttgcacagtagtcaaggtggtccaaagatattggggattgatttgaacttga 425
QY 2187 ttgtgcatagagccccaagctcagagacaggttagagagatgaatagaaacttaaga 2246
Db 426 ttgtgcatagagccccaagctcagagacaggttagagagatgaatagaaacttaaga 485
QY 2247 gaccttaccnaattgaccacagagactgcatlaaigtatg 2291
Db 486 gaccttaccnaattgaccacagagactgcatlaaigtatg 530

RESULT 6
US-09-376-781-7
; Sequence 7, Application US/09376781


```

: Patent No. 6261806
: GENERAL INFORMATION:
: APPLICANT: Banerjee, Papia T.
: APPLICANT: Pallence, Clive
: APPLICANT: Andersson, Goran K.
: TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
: Patent No. 6261806
: TITLE OF INVENTION: Use
: FILE REFERENCE: 61750-267
: CURRENT APPLICATION NUMBER: US/09/376,781
: CURRENT FILING DATE: 1999-08-18
: EARLIER APPLICATION NUMBER: 60/097,015
: EARLIER FILING DATE: 1998-08-18
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7
: LENGTH: 445
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
: OTHER INFORMATION: sequence of a portion of the PPRV-A env gene
: OTHER INFORMATION: corresponding to nucleotides 354-809 of GenBank
: OTHER INFORMATION: NO. 6261806 Y12238.
: US-09-376-781-7

```

Query Match	8.0%;	Score 265;	DB 4;	Length 445;
Best Local Similarity	99.5%;	Pred. NO. 6.5e-107;		
Matches 435; Conservative	0;	Mismatches 1;	Indels 1;	Gaps 1;

QY	2823	ccctaacccttactctctactcgttgtaacttactgaaccggtacaaggtattaatat	2887
Db	1	ccctaacccttactctctactcgttgtaacttactgaaccggtacaaggtattaatat	60
QY	2883	acagactcaaggaggagctcccttggggaccctgtgtgcctgaatatatgtctgcctc	2947
Db	61	acagactcaaggaggagctcccttggggaccctgtgtgcctgaatatatgtctgcctc	120
QY	2943	gataagtaatcccttggtcttaagaccagccacacccccgatgtactccgtgtctag	3007
Db	121	gataagtaatcccttggtcttaagaccagccacacccccgatgtactccgtgtctag	180
QY	3003	ggtttacggttccccaagccccccaataatgaagaatattgtgaaatcctccagatt	3063
Db	181	ggtttacggttccccaagccccccaataatgaagaatattgtgaaatcctccagatt	240
QY	3063	tcctttgcaagcaatvgagctgcataactcttaatgty-ggaattggaatlygccaac	3121
Db	241	tcctttgcaagcaatvgagctgcataactcttaatgty-ggaattggaatlygccaac	300
QY	3122	tctcagaagaacaggaagaatttactcttttgttaacaactccacacagttataactaatt	3181
Db	301	tctcagaagaacaggaagaatttactcttttgttaacaactccacacagttataactaatt	360
QY	3182	aattatgccaatggaatgynaagaatttggcacaacgcygttacaanaagaatlycgaat	3241
Db	361	aattatgccaatggaatgynaagaatttggcacaacgcygttacaanaagaatlycgaat	420
QY	3242	aagcaataaagctgtca 3258	
Db	421	aagcaataaagctgtca 437	

RESULT 7
US-09-075-272-1
; Sequence 1, Application US/09075272
; Patent No. 6136598

GENERAL INFORMATION:
APPLICANT: MILLER, A. DUSTY
APPLICANT: WOLSMOT, GREG
APPLICANT: BONHAM, LYNN
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	1.1%;	Score 35;	DB 3;	Length 8655;
Best Local Similarity	100.0%;	Pred. No. 2.9e-06;		
Matches 35;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

	Y	2203	caaaagctcagacacgacgaagaatgaatacgaac	2237
	I			
Dn	5249	CAAGCTCAGACACGATGAGAGAGATGAATTAGAAC	5283	
	RESULT	8		
	US-09-352-990-7			
	; Sequence 7, Application US/09352990			
	; Patent No. 6255090			
	GENERAL INFORMATION:			
	APPLICANT: Famodu, Layo O.			
	APPLICANT: Orozco, Buddy			
	APPLICANT: Katsalski, Antoni			
	TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetases			
	FILE REFERENCE: BB-1191			
	CURRENT APPLICATION NUMBER: US/09/352,990			
	CURRENT FILING DATE: 1999-07-14			
	EARLIER APPLICATION NUMBER: 60/092,866			
	EARLIER FILING DATE: July 15, 1998			
	NUMBER OF SEQ ID NOS: 29			
	SOFTWARE: Microsoft Office 97			
	SEQ ID NO 7			
	LENGTH: 836			
	TYPE: DNA			
	ORGANISM: Triticum aestivum			
	US-09-352-990-7			

Query Match	0.9%;	Score 30;	DB 4;	Length 836;
Best Local Similarity	100.0%;	Pred. No. 0.00053;		
Matches 30;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 3291 tcactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 742 tcactaaaaaaaaaaaaaaaaaaaaa 771

RESULT 9

US-09-277-716-31
; Sequence 31, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (Insert B)
US-09-277-716-31

Query Match 0.9%; Score 30; DB 4; Length 1279;
Best Local Similarity 100.0%; Pred. No. 0.00051;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3291 tcactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 1237 tcactaaaaaaaaaaaaaaaaaaaaa 1266

RESULT 10

US-08-798-000-3
; Sequence 3, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,000
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,564
; FILING DATE: 13-FEB-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-002010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-798-000-3

Query Match 0.9%; Score 29; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2185 catgtgcatacagaccaccaagctcagg 2213
|||||
Db 96 CATGTGCATACAGACCCCAAGCTCAGG 124

RESULT 11

US-09-248-335-27
; Sequence 27, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 27
; LENGTH: 1198
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-27

Query Match 0.9%; Score 29; DB 3; Length 1198;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3292 cactaaaaaaaaaaaaaaaaaaaaa 3320
|||||
Db 1109 cactaaaaaaaaaaaaaaaaaaaaa 1137

RESULT 12

US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,000
FILING DATE: 12-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,564
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14538A-002010
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-798-000-1

Query Match 0.9%; Score 29; DB 1; Length 1234;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2185 catctgcatagagaccaccaagctcag 2213
|||||
DB 96 CATGTGATACAGACCCCAAGCTCAGC 124

RESULT 13
US-09-232-200-66
Sequence 66, Application US/09232200A
Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MB
CURRENT APPLICATION NUMBER: US/09/232,200A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 2338
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-200-66

Query Match 0.9%; Score 29; DB 4; Length 2338;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3292 cactaataaaaaaaaaaaaaaaaaa 3320
|||||
DB 2295 cactaataaaaaaaaaaaaaaaaaa 2323

RESULT 14
US-09-232-197-66
Sequence 66, Application US/09232197A
Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Gimeno, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21p3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
CURRENT FILING DATE: 1999-01-14
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 66
LENGTH: 2338
TYPE: DNA
ORGANISM: Mus musculus
US-09-232-197-66

Query Match 0.9%; Score 29; DB 4; Length 2338;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3292 cactaataaaaaaaaaaaaaaaaaa 3320
|||||
DB 2295 cactaataaaaaaaaaaaaaaaaaa 2323

RESULT 15
US-09-011-745-3
Sequence 3, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cosset, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 7308
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

Query Match 0.9%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2185 cattgtgcatcacagaccaccaagctcag 2213
|||||
Db 6178 cattgtgcatcacagaccaccaagctcag 6206
|||||

Search completed: February 24, 2002, 07:23:41
Job time: 16030 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 07:23:41 ; Search time 197.25 Seconds

(without alignments)
9425.389 Million cell updates/sec

Title: US-09-171-553B-3

Perfect score: 8209

Sequence: 1 gtgtgtgtacactgtggtgc.....aaaaaaaaaaaaaaaaaaaaa 8209

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2967	36.1	7333	4 US-08-766-528-2	Sequence 2, Appli
2	1466	17.9	3482	3 US-09-111-085-3	Sequence 3, Appli
3	704	8.6	8132	4 US-08-766-528-3	Sequence 3, Appli
4	457	5.6	2462	3 US-09-111-085-1	Sequence 1, Appli
5	375	4.6	8060	4 US-08-766-528-1	Sequence 1, Appli
6	278	3.4	278	4 US-09-376-781-10	Sequence 10, Appli
7	86	1.0	276	4 US-09-376-781-8	Sequence 8, Appli
8	35	0.4	8655	3 US-09-075-272-1	Sequence 1, Appli
9	29	0.4	735	1 US-08-798-000-3	Sequence 1, Appli
10	29	0.4	1234	1 US-08-798-000-1	Sequence 1, Appli
11	29	0.4	7308	4 US-09-011-745-3	Sequence 3, Appli
12	29	0.4	7308	4 US-09-011-745-4	Sequence 4, Appli
13	29	0.4	7616	4 US-09-011-745-2	Sequence 2, Appli
14	29	0.4	8302	1 US-08-258-420-13	Sequence 13, Appli
15	29	0.4	8332	3 US-08-850-961-1	Sequence 1, Appli
16	28	0.3	698	4 US-09-376-781-18	Sequence 18, Appli
17	28	0.3	949	5 PCT-US91-08254-1	Sequence 1, Appli
18	28	0.3	949	5 PCT-US91-08254-2	Sequence 2, Appli
19	28	0.3	1493	4 US-09-376-781-24	Sequence 24, Appli
20	28	0.3	1736	4 US-09-360-197-13	Sequence 13, Appli
21	28	0.3	2000	4 US-09-376-781-25	Sequence 25, Appli
22	28	0.3	2000	4 US-09-376-781-30	Sequence 30, Appli
23	27	0.3	1587	3 US-09-108-020-11	Sequence 11, Appli
24	26	0.3	843	2 US-08-831-570-2	Sequence 2, Appli
25	26	0.3	1098	3 US-09-248-335-35	Sequence 35, Appli
26	26	0.3	1275	3 US-08-725-532A-2	Sequence 2, Appli
27	26	0.3	4032	1 US-08-126-587C-8	Sequence 8, Appli

28	26	0.3	80595	4 US-09-078-294-3	Sequence 3, Appli
29	25	0.3	25	4 US-09-376-781-23	Sequence 23, Appli
30	25	0.3	945	4 US-09-268-364-3	Sequence 3, Appli
31	25	0.3	1000	4 US-09-357-251-19	Sequence 19, Appli
32	25	0.3	1001	1 US-08-728-259A-10	Sequence 10, Appli
33	25	0.3	1001	2 US-08-473-486-10	Sequence 10, Appli
34	25	0.3	1026	2 US-08-713-000-6	Sequence 6, Appli
35	25	0.3	1026	2 US-08-975-316-6	Sequence 6, Appli
36	25	0.3	1026	4 US-09-211-710-6	Sequence 6, Appli
37	25	0.3	1055	1 US-08-702-344-18	Sequence 18, Appli
38	25	0.3	1065	2 US-08-605-106-3	Sequence 3, Appli
39	25	0.3	1622	4 US-09-334-601-9	Sequence 9, Appli
40	25	0.3	1737	4 US-09-416-050A-7	Sequence 7, Appli
41	25	0.3	1737	4 US-09-664-800-7	Sequence 7, Appli
42	25	0.3	1737	4 US-09-665-309-7	Sequence 7, Appli
43	25	0.3	1737	4 US-09-661-569-7	Sequence 7, Appli
44	25	0.3	1826	2 US-08-808-931-15	Sequence 15, Appli
45	25	0.3	1826	3 US-08-808-931-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-08-766-528-2
: Sequence 2, Application US/08766528
: Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF INVENTION: AND METHODS OF USE

CORRESPONDENCE ADDRESS: 74

ADDRESS: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MGP-038CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 7333 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-766-528-2

Query Match 36.1%, Score 2967, DB 4, Length 7333;
Best Local Similarity 99.8%, Pred. No. 0;
Matches 3507; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 1 gtgtgtacagatgtggcccccagcgcttggaataaacctctgtgcgttcgctc 60
|||||

Db 11 GTGTTGTAAGACGTGGGCCCCAGGCGCTTGGAATAAATCCCTTGGTTTGATC 70
Qy 61 aagaccgttcgtgagatgattggggtgcgcctctccagaccggagaggat 120
Db 71 AAGACGGCTCTTGTGATGATTTGGGTGCGCTTCCTCCGACCCGGAGAGGGGAT 130
Qy 121 tgttccttactgacctcattcgtgltggtccgggaaatctcgcacacctta 180
Db 131 TGTTCCTTACTGGGCTTTCATTTGGTGGCTGGGGAATCTTCGACACCCCTTA 190
Qy 181 caaccgagacgaacttggagataaaggatccctcttggaaagtgtgtgtcgcc 240
Db 191 CACCCGGAACCGACTTGGAGTAAGGATCCCTTTGGAACATATGTGTGCGGCC 250
Qy 241 ggcgtctctgtgagatgtctgttccgtgtatgcgcgtcttcgttgcagtctc 300
Db 251 GGGCTCTCTGTCTGAGTGTCTGTTCGGTGATGCGGCTTCGGTTTCAGTGTCTC 310
Qy 301 ctccagaccgttaagagacttggagactgtgtatcagcagactgtctaggagatacaaggctg 360
Db 311 CTCAGACCGTAAGGACTGGAGAGACTGTGATCAGACAGTGTGATGAGAGATCAAGGCTG 370
Qy 361 ccaaccctggggagacgcccgggagggtgggagagccaggagcgctgtgtctctact 420
Db 371 CCACCTTGGGGAGCGCCCGGAGGTGGGAGAGACCAGGAGCGCTGTGCTCCACT 430
Qy 421 gtccgttaagagagaccaggtctgttfttgaagcgaagactccctcccgcgccgtc 480
Db 431 GTCCGTAGAGAGACCGAGTCTGTGTGTAAGGAGAAAGCTTCCCTCCCGCCGCTCCG 490
Qy 481 actcttctgctcgtgtgtgaaagacgagcgagtcggtgtgtctgtatctgtgtt 540
Db 491 ACTCTTTTGGCTGCTGTGGAAGACGGGAGCGGTCGTGTGTGATCTTGTGT 550
Qy 541 ctgtctgtgtgtcttctgtctgtgtgtctgtctgtctgtctgtctgtctgtct 600
Db 551 CTGTTCTGT 610
Qy 601 tgaactaccccccttaagttaactcgaacatctgaactgaactgaactgaactgaact 660
Db 611 TGACTACCCCCCTTAAGTTTACTCTGACCATTTGACTGAGTGAATTAATCCAGGGCTCAT 670
Qy 661 atttgtcagttcaggttaagaaggaacttgcagacttctgtcctctgaatgtgcaa 720
Db 671 ATTGTGACAGTTAGGTTAAGAGGACCTTGGCAGACTTCTGTGCTGTGATGGCCAA 730
Qy 721 cattcagatgtgagatgagacacagagaggaacttaattcctgaattatctgtgtt 780
Db 731 CATTCGATGTTGGATGGCAATCAGAGGGGACCTTTAATTCTGMAATTATCCGGCTGTTA 790
Qy 781 aggaacatcatcttccagactgacccgggtcctcactcgtatcagagagccctatctcta 840
Db 791 AGGCATCATTTTTCAGACTGGACCCGGCTCTCATCTTATCAAGGAGCCCTATATCTTTA 850
Qy 841 cgtgtgcaagatttgcagagaagatcctcgcgcatalggtttaaccatgtctaaataaccaa 900
Db 851 CGTGGCAAGATTTGGCAAGAGATCTCCGCCATGGGTAAACATGGCTAAATTAACCA 910
Qy 901 gaaagcagagttcccggaatccctgtgtcttggagagaaacaaacactgcgcgaaag 960
Db 911 GAAAGCAAGGTCCCGAATCTGGCTTGTGAGAGAAAAACAACACTGCGCGAAAAAG 970
Qy 961 tcgagccctctctcgtatctaccccgagatcgaagagcccgagacttggccgaaacc 1020
Db 971 TCGAGCCCTC-TCTCTGTATCTACCCCGAGATCGAGAGCCCGCACTTGGCGGAACCC 1029
Qy 1021 caactgttcccccaccccttatacagacagaggtgtgtgtgagggga-cctctgcct 1079
Db 1030 CAACCTGTTCCCGACCCCTTATCCAGACAGGGGTGCTGAGGGGACCCCTCTGCCCT 1089
Qy 1080 cctgtgagtcctgggtgtgagagactgtgtgcggagactcgaagccgagagagcgcaac 1139
Db 1090 CCTGAGGCTCCGGTGGTGGAGGAGACTGCTCCGGGAGACTCGAGCCGAGAGAGGCCACCC 1149
Qy 1140 ccgagagcgagacagagatcgcgatalaccgcttgcacactatgtccctcccatgcca 1199
Db 1150 CCGAGCGGAGACAGAGATCGGATATTATCCGCTCGCAGCCATGGCCCTCCATGCCA 1209
Qy 1200 gggggccaattgagccctccagatgtgccccttctctcgcagatctcataattg 1259
Db 1210 GGGGGCCAATTGGAGCCCTCCAGATGTTGGCCCTTTTCTTGAGATCTTATATGG 1269
Qy 1260 aaactaacatccctctctcgcagagatccccaacgctcagaggttggtaggtcc 1319
Db 1270 AAACATACATCCCTCTTCTCGAGAGATCCCAACGCTCAAGGGGTTGGTAGGTCC 1329
Qy 1320 ctatgtctctcaccagacttctgtgagatgtgtlcaaacagctgtcgaagactctc 1379
Db 1330 CTATGTCTCTACACAGCTACTTGATGATGATTTGCAACACACTCTGAGACATCTTC 1389
Qy 1380 acaacgagagcgagagagaaattcgtttagaggtcagaagaaatgttccctgtggcgac 1439
Db 1390 ACAACGAGAGAGCGAGAGAAATCTGTAGAGGCTTAAAMAAAATGTTCTTGGGCGCAG 1449
Qy 1440 gggcgacccacgcaqtlgtcaaaatagatltgacatgagattcccttaccctgcgcggt 1499
Db 1450 GGGCGACCCACCGAGTTCGAAATAGATTTGATGATGGATTTCTTACCTGACCTGCCGCT 1509
Qy 1500 tgggactaacacacggttgaaggttagggagacttgaatcatalcgcacagctctgtg 1559
Db 1510 TGGGACTAACACAGGCTGAAAGGTAGGAGACTTGAAMATCTATCCGACGCTCTGTG 1569
Qy 1560 gcggtctctcggggcgccctcaagagcgccctaaattgtctgaagttagagaggtgtg 1619
Db 1570 GCGGCTCTCGGGGGGCTCAAGAGCGCCCAATATTTGGCTTAAGTGAAGAGGTGATG 1629
Qy 1620 cagggacggaagaaactccctcgttattctcttgaagagctcatalgaagcctcagcg 1679
Db 1630 CAGGACCGAAGAACCTCCCTCGGTATTCTGTGAGAGCTCATGGAAGCCTTCAGCGGG 1689
Qy 1680 ttcaaccttctgtactcactcagagcccaagaagcctcagltgcccgtgctcaat 1739
Db 1690 TTCACCCCTTTGATCTTACCTCAGAGGCCCAAGAACCTCAGTGGCCCTGCTTCAT 1749
Qy 1740 gggcgctcggctctgtatatacaggaagaaacttcaagagctggaaggtttcaagagct 1799
Db 1750 GGGCGCTGGCTCTTGATATCAGAGAAACCTTCAGAACTGGAAGGGTTACAGAGGCT 1809
Qy 1800 gagttaagttatctagtgagagagcgagaaaggtgtltaacaagaggaagacagag 1859
Db 1810 GAGTTACGTGATCTGTGAGAGAGGACAGAAAGGTGTTACAGAAAGGACACAGAGAG 1869
Qy 1860 gagaagagaaagagaaagaaagagagagaaagagagagagagagagagagagag 1919
Db 1870 GAGAAAGAAAGAGAAAGAAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
Qy 1920 caaggaagaaatttgacaaagatttggccgcaatgttgaaggaagagcagagagag 1979
Db 1930 CAAGAGAAATTTGACTTAAGATTTGGCCGCACTGTGTGAAGGAGAGACAGAGAGAG 1989
Qy 1980 agagagagaaattttagaanaattaggtlcaaggccctagacagltcaaggaaacttggcaat 2039
Db 1990 AGAGAGAGATTTTAGAANAATTAGGTACAGGGCCCTGAGACGTAGAGGAACCTGGCAAT 2049
Qy 2040 agagcccatcctcgaagagacaggtgtgtatattgttaagaaagaaagacacttggcaag 2099
Db 2050 AGGACCCCACTCGAAGAGACCACTGTGCGTATTGAAAGAAAAAGACACTGGCGAAG 2109
Qy 2100 aactgcccagaagaggaagaaagaaagagcggaaggtccctaacttagaagaaataagat 2159
Db 2110 AACTGCCCAAGAGAGAGAAAGAAAGACCGAAGGCTCTTACTCTTAGAAGATTAAGAT 2169
Qy 2160 taaggagagagaggttctgcagccctcccgagcccgaggtlaacttgaaggttgaaggg 2219
Db 2170 TAGGGAGAGAGGGGTTGGAACCCCTCCCGAGGCCGAGGGTAACTTTGAAGGTGAGAGGG 2229

QY 2220 caaccagttgagttccttggtatgaacggagccggagcattcagtgctgctcaacacatta 2279
 Db 2230 CAACCGATTGAGTTCTGGTGTATACCGGAGGGAGACANTGCTGCTCAACACATTATA 2289
 QY 2280 ggaataacataaagaaaaaataccttggtgtagtgggtccagaaggcaacggcagctatcca 2339
 Db 2290 GGAATAACTATAAGAAAAAATCTGGTGTATGGGTGCCACAGGGCCAAACGGCAGTATCCA 2349
 QY 2340 tggactacccgaagaacccgttgacttggtggtggtggtggtggtggtggtggtggtggtc 2399
 Db 2350 TGGACTACCCGAGAAACCGGTGACTTGGAGTGGAGGAGGAGTAAACCCACTGTTTCTGGTGC 2409
 QY 2400 atcccttaagtgcccgagtaacccctctcagtagagacttaactgacccaagaatggagctcaa 2459
 Db 2410 ATCCCTTAGTGCCCGAGTAAACCCCTTCTAGTAGAGACTTACTGACCAAGATGGAGCTCAAA 2469
 QY 2460 attctcttgaacaagaagaacagaagtgctgctgtaataaacaacccctcagctgtgtgt 2519
 Db 2470 ATTCTTCTTGAACAAGAGAAAGACCGAGAGTGCTGTGATATACAAACCCACTGCTGTGTGT 2529
 QY 2520 accctcaattagatgtagaataatcagataatctccccaagtaagcctgatacaagat 2579
 Db 2530 ACCCTCAATTAGATGATGATATCGACTATATCTCCCAAGTAAAGCCTGATCAAGAT 2589
 QY 2580 atacagctcgtgtgagagcagttccccaagcctggagcagaacccgaaggatggtgtgt 2639
 Db 2590 ATACAGCTCCTGTTGAGACAGTATCCCAAGCCTGGGCAAAACCCGAGGAGTGGGTGTGG 2649
 QY 2640 gcaaaagcaagtctccccaagcagttatcaactgaagcagtgctacacagatcagtc 2699
 Db 2650 GCAAAAGCAAGTCTCCCAAGCAGTATATCACTGAAGCCCACTGCTACACAGATATACGTC 2709
 QY 2700 agacagtaaccccttgagtagagaggtctcgagaagaaatltggccgcaatgttcaagaatta 2759
 Db 2710 AAGACGTAACCCCTTGAGTAGAGAGGCTCGAGAAGAAATTTGGCCGATGTTCAAGATTA 2769
 QY 2760 atcccaagggagcattcctgctgtcctcctcctcctcctcctcctcctcctcctcctcctcct 2819
 Db 2770 ATCCCAAGGGGAGCATTCTGAGTGTCTGTCATATCCCTTGGATATCTCCCTGCTACCGGTT 2829
 QY 2820 aggaagccttgaggcaatgatattatcagcagtagacagatcgtgagagagatcaataaaga 2879
 Db 2830 AGGAGCCTTGAGGCAATGATATATGACCAAGTAGAGAGGAGTGTGAGAGAGTCAATATAAGG 2889
 QY 2880 gtgcagagacataacccaacggtcccggaacccctataacccctctgagagccctccgct 2939
 Db 2890 GTGCAGAGACATACACCCAGGCTCCCGAACCCTTATTAACCTCTTGAGGCGCTCCCGCT 2949
 QY 2940 gaacggaaactgtgacagatattgacttaaaagaatgcctcctcctcctcctcctcctcctcctc 2999
 Db 2950 GAACGGAACTGTTACAGATATTGGACTTAAAGATGCTTCTTCTGCTGAGATTACAC 3009
 QY 3000 cccaactagccaacacatttctgctcctcgaaatggagagatccaagttaagggaagaacggg 3059
 Db 3010 CCCACTAGCCAAACACATTTTACCTTCGATGAGAGAGATCCAGGTACGGGAGAACCCGGG 3069
 QY 3060 cagactcaacttgagcccgactgcccgaagggttcaagaactcccgacacatccttgaagaa 3119
 Db 3070 CAGCTCACTGAGCCGACCGACGCCCCAAGGGTTCAAGACTCCCGACACATCTTTAGCAGA 3129
 QY 3120 gacctacacagagacctgagccaactcagaatccaacacccctcaagtgagcctcctcag 3179
 Db 3130 GCCTTACACAGGGAGCGGCGCACTTCAGAGATCCAAACACCTCTAGGTGACCTCTCTCAG 3189
 QY 3180 taagtggatgagcctgtctctggtgggagccacccaacagagactgcttaaaagtagcagaag 3239
 Db 3190 TACCTGGATGACCTGCTTCTGGGGGGAGCCACCAACAGGACTGCTTAAAGGTACGAAG 3249
 QY 3240 gcaactactgctggaattgtctgacactagctacagagcctctgtctaagaaggccagatt 3299
 Db 3250 GCACACTAGCTGGAATTGTCTGACACTAGCTACAGAGCCCTCTCTAAGAAAGGCCAGATT 3309
 QY 3300 tgcagagagaggttaacataactctgggttacaagtttgcggggcgaggcagatgctgacgc 3359

Db 3310 TGCAGAGAGAGAGTATACATACTTGGGCTACACTTTCGCGGGCGGACAGATGCTGACG 3369
 QY 3360 gaagcagcagaagaacactgtagtcaagataccggccccaacacagacccaagaatgaga 3419
 Db 3370 GAGGACAGCGAAGAAACCTGATGTCAGATACCGGCGCCCAACACAGCCAAACAGTGAAG 3429
 QY 3420 gagtlttgggagacactgatttgcagacgtgacccggggttggcagacttaaga 3479
 Db 3430 GAGTTTGGGAGACAGCTGATTTTGCAGACTGTGATCCCGGGGTTTGCAGCTTACGA 3489
 QY 3480 gccccactcaccgcttaacccaagaagaagggg 3514
 Db 3490 GCCCCACTTACCCGCTAACCAAGAAAGAGGGG 3524

RESULT 2
 US-09-111-085-3
 : Sequence 3, Application US/09111085
 : Patent No. 6100034
 : GENERAL INFORMATION:
 : APPLICANT: Stoye, Jonathan P
 : APPLICANT: Weiss, Robin A
 : TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
 : FILE REFERENCE: 4238/75168
 : CURRENT APPLICATION NUMBER: US/09/111,085
 : EARLIER FILING DATE: 1998-07-07
 : EARLIER APPLICATION NUMBER: GB 9710154.7
 : NUMBER OF SEQ ID NOS: 16
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO: 3
 : LENGTH: 3482
 : TYPE: DNA
 : ORGANISM: Porcine retrovirus
 US-09-111-085-3

Query Match 17.9%; Score 1466; DB 3; Length 3482;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 2536; Conservative 0; Mismatches 13; Indels 4; Gaps 3;
 QY 5639 ggcgcacccctccgactcgggttgagagccggaagaagatlagatcccttaagcttcg 5698
 Db 930 ggcgcacccctccgactcgggttgagagccggaagaagatlagatcccttaagcttcg 989
 QY 5699 cctccatgctggttccctactctaaacataactccccaagcagtagtaaacgctta 5758
 Db 990 cctccatgctggttccctactctaaacataactccccaagcagtagtaaacgctta 1049
 QY 5759 tagacagctlogaacccccataacacttaccctacttgctgatttataacctgata 5818
 Db 1050 tagacagctlogaacccccataacacttaccctacttgctgatttataacctgata 1109
 QY 5819 cgggtgtcactgtraaataacactcagagtggtgtcctcctcctcctcctcctcctcctcctc 5878
 Db 1110 cgggtgtcactgtraaataacactcagagtggtgtcctcctcctcctcctcctcctcctcctc 1169
 QY 5879 tgaattctcctccgattgataaccccgctglttaaaagccacactcccaactagctc 5938
 Db 1170 tgaattctcctccgattgataaccccgctglttaaaagccacactcccaactagctc 1229
 QY 5939 gtagttatggttctatcttgctccagagcagagagaagaataactgtgaggggttctg 5998
 Db 1230 gtagttatggttctatcttgctccagagcagagagaagaataactgtgaggggttctg 1289
 QY 5999 ggaatacctctgtagagatgagctgctcaactcccaacagatgagagctggaatggc 6058
 Db 1290 ggaatacctctgtagagatgagctgctcaactcccaacagatgagagctggaatggc 1349
 QY 6059 cgactctctccagagacgggttaaatctccttctcaattccggcccgaggcaagtaca 6118

Db 1350 cgaatctctccagagccgggttaaaatctctcttctgaaltccgcccggcgaatgaca 1409
 QY 6119 aatgtatgaactatataaagaataagagctgctcccatcagacttgaatataaga 6178
 Db 1410 aagtgatgaactatataaagaataagagctgctcccatcagacttgaatataaga 1469
 QY 6179 taagttaacatg - aagaagaagaagaagaataatcacaagtgataatgtatgact 6236
 Db 1470 taagttaacatgagaagaagaagaagaataatcacaagtgataatgtatgact 1529
 QY 6237 ggggaatgattttat - tataatgaggggagagaggttcacattcaacattcgacta 6295
 Db 1530 ggggaatgattttatataaataatgaggggagagaggttcacattcaacattcgacta 1589
 QY 6296 ggaatagaacgggag 6355
 Db 1590 ggaatagaacgggag 1649
 QY 6356 agggag 6415
 Db 1650 agggag 1709
 QY 6416 ggcctgacataaacaagc 6475
 Db 1710 ggcctgacataaacaagc 1769
 QY 6476 ctagaactcccaaggtgttctctgttgaagaagaagaagaagaagaagaagaagaagaaga 6535
 Db 1770 ctagaactcccaaggtgttctctgttgaagaagaagaagaagaagaagaagaagaagaaga 1829
 QY 6536 gaacttcccaagcacaactcccaagcacaactcccaagcacaactcccaagcacaactccca 6595
 Db 1830 gaacttcccaagcacaactcccaagcacaactcccaagcacaactcccaagcacaactccca 1889
 QY 6596 talccctcagggcctctcttcttcttcttcttcttcttcttcttcttcttcttcttctt 6655
 Db 1890 talccctcagggcctctcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1949
 QY 6656 aagaagatagaataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 6715
 Db 1950 aagaagatagaataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2009
 QY 6716 ccgggaaggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6775
 Db 2010 ccgggaaggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2069
 QY 6776 ctgtgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 6835
 Db 2070 ctgtgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 2129
 QY 6836 gggcagtgcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 6895
 Db 2130 gggcagtgcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2189
 QY 6896 attctgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 6955
 Db 2190 attctgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 2249
 QY 6956 ttgataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 7015
 Db 2250 ttgataatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2309
 QY 7016 ctgtaatgtcgcgattaggaag 7075
 Db 2310 ctgtaatgtcgcgattaggaag 2369
 QY 7076 caagagacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 7135
 Db 2370 caagagacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2429
 QY 7136 tccgagccttaagaag 7195
 Db 2436 tccgagccttaagaag 2489

QY 7196 tgggttctacagaacccggag 7255
 Db 2490 tgggttctacagaacccggag 2549
 QY 7256 cagccttaagaagaagaatgt 7315
 Db 2550 cagccttaagaagaagaatgt 2609
 QY 7316 tgaacaagccttaagaagaagaatgt 7375
 Db 2610 tgaacaagccttaagaagaagaatgt 2669
 QY 7376 ggttgaagaagaatgt 7435
 Db 2670 ggttgaagaagaatgt 2729
 QY 7436 ggcctcagtagtgcct 7495
 Db 2730 ggcctcagtagtgcct 2789
 QY 7496 ttgccttctgttagagaag 7555
 Db 2790 ttgccttctgttagagaag 2849
 QY 7556 aagccttctgttagagaag 7615
 Db 2850 aagccttctgttagagaag 2909
 QY 7616 attaacaagaagaagaatgt 7675
 Db 2910 attaacaagaagaagaatgt 2969
 QY 7676 ccaggaaggttaataaagaatgt 7735
 Db 2970 ccaggaaggttaataaagaatgt 3029
 QY 7736 aatagtgatgaag 7795
 Db 3030 aatagtgatgaag 3089
 QY 7796 ataaagaagaagaatgt 7855
 Db 3090 ataaagaagaagaatgt 3149
 QY 7856 agaagaatgt 7915
 Db 3150 agaagaatgt 3209
 QY 7916 tgcgaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7975
 Db 3210 tgcgaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3269
 QY 7976 agcgcggcctctcgaatgt 8034
 Db 3270 agcgcggcctctcgaatgt 3329
 QY 8035 taagaatgt 8094
 Db 3330 taagaatgt 3389
 QY 8095 tccgacatcgggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 8154
 Db 3390 tccgacatcgggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 3449
 QY 8155 gcttgaataataaataatct 8217
 Db 3450 gcttgaataataaataatct 3482


```
Db 4470 GACCGGGTGGCCAAAGCCAGCCAGGGTGTAACTTCTGCTTAAATAAGAAATGCC 4529
Qy 4533 aaagcccaagaccagacagacag lacaccctagaagactgagcaagataaaagata 4592
Db 4530 AAAGCCCGAGAACCCAGACGACATACACCTACAAACATGCGCAAGAGATAAAAAGATA 4589
Qy 4593 gaccagtctctgagactccgagggagactgtataccctaatatggaaagaatccgt 4652
Db 4590 GACCAAGTCTCTGAGACTCCGGAAGGAGCCTGCTATACATGAGTGGAGGAATCTCTG 4649
Qy 4653 ccccaacaagaagggttagaatatgtccacaagatacatgtctaacccactagaact 4712
Db 4650 CCCACAAAGAGGGTTAGATATGTCAACAGATACATGCTTAACCCACTAGGAGACT 4709
Qy 4713 aaacactgagcaggtgtgacagacalccctatcatgtctcgaagctaccagagtg 4772
Db 4710 AAACACCTGCAGCAGTGTGTACAAACATCCCTTATCATCTTGTGAGGCTACAGAGAGTG 4769
Qy 4773 gctgactcgtgtgcaaacattgtgacctgcaagctgttaagttaactccctcaga 4832
Db 4770 GCTGACTCGTGTGCAAACTTGTGTGCTCCCTGCAGCTGCTTAAATGCTTAACTCCACGA 4829
Qy 4833 ataccctcaagaaagacacaaagggaagccacccagcgctgactgggaagtgagctc 4892
Db 4830 ATGCTCTCAGGGAAGACACTAAGGGAAGCACCAGGCTCTACTGGAGTGGACTTC 4889
Qy 4893 actgaggtaaagccggttaataacggaacaaatatactatgtgttttgtaagacatt 4952
Db 4890 ACTGAGGTAAAGCCGGCTTAATACGGAACAAATACCTATTGTGTTTGTGAGACACTTC 4949
Qy 4953 tcagagtggttagagcttaactactactaagaagaagactcaacggtgtgtgaagaa 5012
Db 4950 TCAGGATGGTAAAGGCTTACTACTAAGAAAGACACTTCAACCTGTGTGCTTAAAAAA 5009
Qy 5013 atactgaggaagaattttccaaagatttgaaataccctaaggttaaggtcagaacaatgt 5072
Db 5010 ATACTGGAAGAATTTTTCACATTTTGAATACCTAAGGTAAAGGCTCAGACAAATGGT 5069
Qy 5073 ccaagcttctgcccaggttaagtcaggagactgcccagaatattgggaattgtagaa 5132
Db 5070 CCAGCTTTTGTGTCGCCAGTAAGTCAAGGACTGGCCAGATATTTGGGATTTGTTGAAA 5129
Qy 5133 ctgcatgtgcatagacagaccacaagctcagagacagtgtagaagatgaatagaacatt 5192
Db 5130 CTGCATTGTGCTACAGACCCCAAGCTCAGAGCAGTAGAGAGATGAATGAACCAATT 5189
Qy 5193 aaagagacccttac 5206
Db 5190 AAAGAGACCCCTTAC 5203

RESULT 4
US-09-111-085-1
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER APPLICATION NUMBER: 1998-07-07
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-1
```

```
Query Match 5.6%; Score 457; DB 3; Length 2462;
Best Local Similarity 99.8%; Pred. No. 4,1e-187;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7335 gtttagagagcgtcgaaggaagaaagagctaccaggggtgtgttgaagatgtgtcaa 7394
Db 1938 gtttagagagcgtcgaaggaagaaagagctaccaggggtgtgttgaagatgtgtcaa 1997
Qy 7395 caggtctcttgatgagacacccctgtctgtcctcgaagggccctcgaagtcgtct 7454
Db 1998 caggtctcttgatgagacacccctgtctgtcctcgaagggccctcgaagtcgtct 2057
Qy 7455 cctgttacttaagttgggcttctgttaataagttgtgtgtgttgaagagc 7514
Db 2058 cctgttacttaagttgggcttctgttaataagttgtgtgtgttgaagagc 2117
Qy 7515 agttagtgagtcacagatctgtacttaggcaacagtaacaaagcctctgaagcagc 7574
Db 2118 agttagtgagtcacagatctgtacttaggcaacagtaacaaagcctctgaagcagc 2177
Qy 7575 agaactgaacctctagccttccagttcttaagattagaacttaacaagaagaagtg 7634
Db 2178 agaactgaacctctagccttccagttcttaagattagaacttaacaagaagaagtg 2237
Qy 7635 gggatgaaagatgaaatgaaccttaacctctccagaaccggaaggttaataaaag 7694
Db 2238 gggatgaaagatgaaatgaaccttaacctctccagaaccggaaggttaataaaag 2297
Qy 7695 ctctaattgccccgaattacagaccctgtcgtcgtcgaagtaaataggtagaagtcaca 7754
Db 2298 ctctaattgccccgaattacagaccctgtcgtcgtcgaagtaaataggtagaagtcaca 2357
Qy 7755 ctctcatgttccaagggccctgtcatcctgtgcccgaagtaagatacagaagaatgtga 7814
Db 2358 ctctcatgttccaagggccctgtcatcctgtgcccgaagtaagatacagaagaatgtga 2417
Qy 7815 cttaatgcctatctgagctctgttaaac 7842
Db 2418 cttaatgcctatctgagctctgttaaac 2445

RESULT 5
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
```

REFERENCE/DOCKET NUMBER: MGP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-1

Query Match 4.6%; Score 375; DB 4; Length 8060;
 Best Local Similarity 98.7%; Pred. No. 6,7e-152;
 Matches 1395; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 2973 gatgcctcttcgctgagattacacccactagccacacacttttgctcgaatg 3032
 |||||||
 Db 5495 GATGCCTTCTTCGCTGAGATTACACCCACTAGCCACACTTTTGGCTTCGATG 5554
 QY 3033 agagatcaggtacggaagaacccgagcctacactgaaacccgaaaggttc 3092
 |||||||
 Db 5555 AGAGATCCAGGTACGGGAACCCGGCAGCTCAGCCGACCTGCCCAAGGGTTC 5614
 QY 3093 aaaaactccccgacatctttgagaaagccctacacagagacttgccaaactcaagatc 3152
 |||||||
 Db 5615 AAGAACTCCCGACCATCTTGAAGAACCCCTACACAGACCTGGCCAACTTCAGATC 5674
 QY 3153 caaacacctcaggtagacccctccacagtagatgacgtcttcgacggaagcacc 3212
 |||||||
 Db 5675 CAACACCTCAGGTGAGACCCCTCCACGATGATGAGTACTGCTTGGGGGAGGACAC 5734
 QY 3213 aaacagagatgcttagaaggtacgaaggaactacgtctggaattctgaactaggtac 3272
 |||||||
 Db 5735 AAACAGGACTGCTTAGAAGGACGAAAGGCACTACTGCTGAATTTGCTGACCTAGGCTAC 5794
 QY 3273 agagccctctgtaagaagggccagatttgagagagaagtaactacttgggtacat 3332
 |||||||
 Db 5795 AGAGCCTCTGTAAGAAAGGCCAGATTTGAGAGAGAGGTAACTATCTTGGGGTACAGT 5854
 QY 3333 ttgagggcgagcagatgctgacgaggaagcaggaagaaactgactcagataccg 3392
 |||||||
 Db 5855 TTAGGGAGGCGGACGATGCTGACGAGGACGACGAGAAACTGTAGTCCGATACCG 5914
 QY 3393 gccccaacacagcacaagaagtgaagagtttttgggagcagctggaatttgacagctg 3452
 |||||||
 Db 5915 GCCCAACACACAGCCAAATGAAGAGAGTATTTGGGGACAGCTGATTTTGCACACTG 5974
 QY 3453 tggatcccgaggttttgagacttagagcccaactctacccgttaacccaagaagaagg 3512
 |||||||
 Db 5975 TGGATCCCGGGGTTTGAGACTTAGAGCCCACTCTACCCGGCTAACCAAGAAAGAGG 6034
 QY 3513 ggaattccggggctccctgagcaccagaagcatttgatgcttcaaaaagccctgctg 3572
 |||||||
 Db 6035 GAATTCCTCGGGCTCTGTGACACAGAAAGCATTTGATGCTATCAAAAAGGCCCTGCTG 6094
 QY 3573 agcgacactgctctgacccctccctgacgtaactaaaccttaaccttlatgtgagatg 3632
 |||||||
 Db 6095 AGCGACACTCTCTGCGCCCTCCCTGACGTAACTAAACCCCTTATGATGTGATGAG 6154
 QY 3633 cgttaaggaagtagcccgaggaagtttaacccaacaccttgagacatggaggaagactgtc 3692
 |||||||
 Db 6155 CGTAAGGAGTAGCCCGGAGATTITTAACCAAAACCTTAGACCATGAGAAAGACCTGTGC 6214
 QY 3693 gcttaactgtcaagaagacttgatcctgttagccaggtgttgccgctatgctgaaagact 3752
 |||||||
 Db 6215 GCTTACTGTCAAGAAGAGTCTGATCTGTAGCAAGTGTGGCCCATATGSCCTGAAGGCT 6274
 QY 3753 atcgacagctggtgcatactgtgtcaaggaagcgtgacaacttgacttggagc-agaatat 3811
 |||||||
 Db 6275 ATCGACACTGTGCGCATACTGTCAAGAGACGCTGACAAATTTGACTTTGGGACAAAGATAT 6334

QY 3812 aactgtaagcccccatgcatltgagaacatgctcggaagccccacagacatgat 3871
 |||||||
 Db 6335 AACTGTAATAGCCCCCATGATTTGGAGAAATCTGTTGGGACAGCCCCACAGCATGTGAT 6394
 QY 3872 gaccacgcccgcagatgacacacatataagagcgtcttcacagagaaggttaacttcgc 3931
 |||||||
 Db 6395 GACCAAGCGCCGATGACCCACTATCAAAAGCGCTGCTTCACAGAGAGGTGACGTTCGC 6454
 QY 3932 tccaccagcgccttcaacctctgcaactctctctgcttgaagaagactgataccagctgac 3991
 |||||||
 Db 6455 TCCACCAACCGCTTCAACCCCTGCCACCTCTTGGCTTAAGAGACTGTGATGAACCACTGAC 6514
 QY 3992 tcatgattgccaactatgatgaggaagactggttccgaaggaaccttaagacat 4051
 |||||||
 Db 6515 TCATGATTGCCATCACTATTGATTTGAGAGACTGGGTCCGACAGACCTTACAGACAT 6574
 QY 4052 accgctgactggaagaagtgtcaacctgttcaactgacggaagcagctatgtgttgaag 4111
 |||||||
 Db 6575 ACCGCTGACTGAGAAAGTGTAACTGCTGACCTGACGGAAGCACCTATGTGTGGAAGG 6634
 QY 4112 taagagatgctgtggtggtgaggtgtgtgaaggaacccgacagatctggtccagagcct 4171
 |||||||
 Db 6635 TAAAGAGATGCTGGGGCGCGCTGTGAGACGGACCCGACGATCTGGCCAGAGCT 6694
 QY 4172 gccggaaggaactcagcgaagaaggtcgaactcattgaccccaagcagcttgcgct 4231
 |||||||
 Db 6695 GCCGGGAGGACTTATGACCAAAAGGCTGAGCTATGGCCCTCAGCAAGCTTTCGGCT 6754
 QY 4232 gccggaaggaatccataaatacattatatacagacagcagatgacttgcagctgaca 4291
 |||||||
 Db 6755 GSCCAAGGGAATCCATTAACATTTATACGAGACAGAGTATGCTTTGGACATGCACA 6814
 QY 4292 cgtacaaggggcatctataaacaaggggtgtgttccctcagaaggaaggaataaa 4351
 |||||||
 Db 6815 CGTACATGGGCGCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGGAATATAA 6874
 QY 4352 gaacaaagagaatcttaagcctatagaaacc 4385
 |||||||
 Db 6875 GAACAAGAGGAATTTCTAGGCTATTAGAACCC 6908

RESULT 6
 US-09-376-781-10
 ; Sequence 10, Application US/09376781
 ; Patent No. 6261806
 ; GENERAL INFORMATION:
 ; APPLICANT: Banerjee, Papia T.
 ; APPLICANT: Patience, Clive
 ; APPLICANT: Andersson, Goran K.
 ; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 ; Patent No. 6261806
 ; FILE OF INVENTION: use
 ; FILE REFERENCE: 61750-267
 ; CURRENT APPLICATION NUMBER: US/09/376,781
 ; CURRENT FILING DATE: 1999-08-18
 ; EARLIER APPLICATION NUMBER: 60/097,015
 ; EARLIER FILING DATE: 1998-08-18
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patentl Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 278
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
 ; OTHER INFORMATION: sequence of part of poEV env gene corresponding to
 ; OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
 US-09-376-781-10

Query Match 3.4%; Score 278; DB 4; Length 278;
 Best Local Similarity 100.0%; Pred. No. 4.1e-110;

	Matches	278:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	6057	gccgattctctcccgagcgccggttaaattctcccttttccaattccgctggcggaagta	6116							
Db	1	gccgattctctcccgagcgccggttaaattctcccttttccaattccgctggcggaagta	60							
QY	6117	caaaatgatgaacctataataagaataagctgtctcccatcagactagattactctaa	6176							
Db	61	caaaatgatgaacctataataagaataagctgtctcccatcagactagattactctaa	120							
QY	6177	gataagttctcactgaagaagaaacgcggagaaatattcaaaagtggtaatgtatgagct	6236							
Db	121	gataagttctcactgaagaagaaacgcggagaaatattcaaaagtggtaatgtatgagct	180							
QY	6237	gggggaatgcttttattatatalatgycgggggagcaggcttcacctttaacatctgccttag	6296							
Db	181	gggggaatgcttttattatatalatgycgggggagcaggcttcacctttaacatctgccttag	240							
QY	6297	gataagagacggggacagacagccccctgttgcaatggagc	6334							
Db	241	gataagagacggggacagacagccccctgttgcaatggagc	278							

;; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,000
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,564
; FILING DATE: 13-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-002010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-798-000-3

Query Match 0.4%; Score 29; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatcacagaccaccaagctcag 5164
Db 96 CATTGTGCATACAGACCCCAAGCTCAGG 124

RESULT 10
US-08-798-000-1
; Sequence 1, Application US/08798000
; Patent No. 5766945
; GENERAL INFORMATION:
; APPLICANT: Miller, A. Dusty
; TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,000
; FILING DATE: 12-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,564
; FILING DATE: 13-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-002010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear

; MOLECULE TYPE: cDNA
US-08-798-000-1

Query Match 0.4%; Score 29; DB 1; Length 1234;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatcacagaccaccaagctcag 5164
Db 96 CATTGTGCATACAGACCCCAAGCTCAGG 124

RESULT 11
US-09-011-745-3
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

Query Match 0.4%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5136 catgtgcatcacagaccaccaagctcag 5164
Db 6178 catgtgcatcacagaccaccaagctcag 6206

RESULT 12
US-09-011-745-4
; Sequence 4, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 7308
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
OTHER INFORMATION: construct
US-09-011-745-4

Query Match 0.4%; Score 29; DB 4; Length 7308;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatagagaccaccaagctcag 5164
|||||
DB 6178 catgtgcatagagaccaccaagctcag 6206

RESULT 13
US-09-011-745-2
Sequence 2, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weists, Robin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cosset, Francois-Ioic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GH96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GH9517263.1
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 7616
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
OTHER INFORMATION: construct
US-09-011-745-2

Query Match 0.4%; Score 29; DB 4; Length 7616;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatagagaccaccaagctcag 5164
|||||
DB 5339 catgtgcatagagaccaccaagctcag 5367

RESULT 14
US-08-258-420-13/C
Sequence 13, Application US/08258420
Patent No. 5710037
GENERAL INFORMATION:
APPLICANT: Nienhuis, Arthur W.
APPLICANT: Vanin, Elio F.
TITLE OF INVENTION: No. 5710037e1 Retroviral Envelope and LTR and Retroviral Vectors
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,420
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Olstein, Elliot M.
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 271010-208
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 8202 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: viral genome
US-08-258-420-13

Query Match 0.4%; Score 29; DB 1; Length 8202;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5136 catgtgcatagagaccaccaagctcag 5164
|||||
DB 2535 catgtgcatagagaccaccaagctcag 2507

RESULT 15
US-08-850-961-1
Sequence 1, Application US/08850961
Patent No. 6013517
GENERAL INFORMATION:
APPLICANT: Respass, James G.
APPLICANT: De Polo, Nicholas J.
APPLICANT: Chada, Sunil
APPLICANT: Sauter, Sybille
APPLICANT: Bodner, Mordechai
APPLICANT: Driver, David A.
TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation, Intellectual Property - R440
STREET: P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,961
FILING DATE: 05-MAY-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kruse, No. 6013517/man J.
REGISTRATION NUMBER: 35,235
REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-3520
TELEFAX: (510) 655-3542

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-850-961-1

```

```

Query Match          0.4%; Score 29; DB 3; Length 8332;
Best local Similarity 100.0%; Pred. No. 0.0036;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5136 catlgtgcatacagaccacaagctcag 5164
   ||||||||||||||||||||||||||||
DB 5232 CATTGTGCATACAGACCCCAAGCTCAGG 5260

```

Search completed: February 24, 2002, 07:30:07
 Job time: 16416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on: February 24, 2002, 07:30:07 ; Search time 197.25 Seconds
        (without alignments)
        3394.012 Million cell updates/sec
```

Title:	US-09-171-553B-9
Perfect score:	2956
Sequence:	1 tgccttttagtgtagaac.....aaaaaaaaaaaaaaaaaa 2956

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

word size : 0

Total number of hits satisfying chosen parameters: 702406

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

post-processing: Listing first 45 summaries

Database :

```
Issueec_Patents_NA:*
1:/cgn2_6/prodcta/2/ina/5A_COMB.seq.*
2:/cgn2_6/prodcta/2/ina/5B_COMB.seq.*
3:/cgn2_6/prodcta/2/ina/6A_COMB.seq.*
4:/cgn2_6/prodcta/2/ina/6B_COMB.seq.*
5:/cgn2_6/prodcta/2/ina/pcrUS_COMB.seq.*
6:/cgn2_6/prodcta/2/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1743	59.0	3482	3	US-09-111-085-3	Sequence 3, Appl1
2	813	27.5	7333	4	US-08-766-528-2	Sequence 2, Appl1
3	457	15.0	2462	3	US-09-111-085-1	Sequence 1, Appl1
4	147	5.0	276	4	US-09-376-781-8	Sequence 8, Appl1
5	137	4.6	278	4	US-09-376-781-10	Sequence 10, Appl1
6	98	3.3	8132	4	US-08-766-528-3	Sequence 3, Appl1
7	91	3.1	8060	4	US-08-766-528-1	Sequence 1, Appl1
8	28	0.9	1493	4	US-09-376-781-4	Sequence 14, Appl1
9	28	0.9	1736	4	US-09-360-197-13	Sequence 13, Appl1
10	28	0.9	2000	4	US-09-376-781-25	Sequence 25, Appl1
11	28	0.9	2000	4	US-09-376-781-30	Sequence 30, Appl1
12	27	0.9	1587	3	US-09-108-020-11	Sequence 11, Appl1
13	26	0.9	843	2	US-08-831-570-2	Sequence 2, Appl1
14	26	0.9	1098	3	US-09-248-335-35	Sequence 35, Appl1
15	26	0.9	1275	3	US-08-725-532A-2	Sequence 2, Appl1
16	26	0.9	4032	1	US-08-126-567C-8	Sequence 8, Appl1
17	26	0.9	80595	4	US-09-078-284-3	Sequence 3, Appl1
18	25	0.8	25	4	US-09-376-781-33	Sequence 23, Appl1
19	25	0.8	698	4	US-09-376-781-18	Sequence 18, Appl1
20	25	0.8	945	4	US-09-268-364-3	Sequence 3, Appl1
21	25	0.8	1000	4	US-09-357-251-19	Sequence 19, Appl1
22	25	0.8	1001	1	US-08-728-259A-10	Sequence 10, Appl1
23	25	0.8	1001	2	US-08-473-486-10	Sequence 10, Appl1
24	25	0.8	1026	2	US-08-713-000-6	Sequence 6, Appl1
25	25	0.8	1026	2	US-08-975-316-6	Sequence 6, Appl1
26	25	0.8	1026	4	US-09-211-710-6	Sequence 6, Appl1
27	25	0.8	1055	1	US-08-702-344-18	Sequence 18, Appl1

ALIGNMENTS

28	2.5	0.8	1066	2	US-08-605-106-3	Sequence 3, Appl1
29	2.5	0.8	1622	4	US-09-334-601-9	Sequence 9, Appl1
30	2.5	0.8	1737	4	US-09-416-050A-7	Sequence 7, Appl1
31	2.5	0.8	1737	4	US-09-664-800-7	Sequence 7, Appl1
32	2.5	0.8	1737	4	US-09-665-309-7	Sequence 7, Appl1
33	2.5	0.8	1737	4	US-09-661-569-7	Sequence 13, Appl1
34	2.5	0.8	1826	2	US-08-808-931-15	Sequence 15, Appl1
35	2.5	0.8	1826	3	US-08-808-323-15	Sequence 15, Appl1
36	2.5	0.8	1826	3	US-09-050-603A-15	Sequence 15, Appl1
37	2.5	0.8	1826	3	US-09-102-420B-15	Sequence 15, Appl1
38	2.5	0.8	1889	3	US-08-861-747-3	Sequence 3, Appl1
39	2.5	0.8	1957	4	US-09-352-990-11	Sequence 11, Appl1
40	2.5	0.8	2158	1	US-07-602-608-1	Sequence 1, Appl1
41	2.5	0.8	2158	1	US-08-261-578-1	Sequence 1, Appl1
42	2.5	0.8	2351	1	US-08-166-091A-1	Sequence 1, Appl1
43	2.5	0.8	3312	4	US-08-697-954-1	Sequence 1, Appl1
44	2.5	0.8	4337	3	US-09-187-049-1	Sequence 1, Appl1
45	2.4	0.8	40	4	US-09-306-290-30	Sequence 30, Appl1

```

RESULT 1
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
; US-09-111-085-3

```

Query Match	59.0%	Score 1743	DB 3	Length 3482
Best Local Similarity	99.3%	Pred. No. 0		
Matches 2913	0	Mismatches 21	Indels 1	Gaps 1

QY	1	tgcttttaaggtcttagaagaaccccccttgagcaagtttggctcgtgaccccccttaagattgctc	60
Db	548	tgcttttttaaggtcttagaagaaccccccttgagcaagtttggctcgtgaccccccttaagattgctc	607
QY	61	acggagggaacccccccgcttggtagaaattgctctctgacagatggcgtgatgctgctt	120
Db	608	acgaggaggacccccccgcttggtagaaattgctctctgacacatagctgctgagtgctgcttc	667
QY	121	cccgagccttctgtctctagagctcaagagcgctcgaagctgggtgtagagcaacgaagctggaagc	180
Db	668	cccgagcctctgtctctctagagctcaagagcgctcgaagctgggtgtagagcaacgaagctggaagc	727
QY	181	agcttcgcgggaagcgctacctcagagagagagagagactctgcaagttccacatcgcctccaaattg	240
Db	728	agcttcgcgggaagcgctaccccaagagagagagagagactctgcaagttccacatcgcctccaaattg	787
QY	241	gagattcagctctatgttgtagaacgcacacgctgtagaggaacactcgcgaactcgtgtggaagagac	300
Db	788	gagattcagctctatgttgtagaacgcacacacgctgtagaggaacactcgcgaactcgtgtggaagagac	847
QY	301	cttatctcgtactctttgagcaacacaaacagctgtggaagatcgaaggaatctccacccctgga	360
Db	848	cttatctcgtactctttgagcaacacaaacagctgtggaagatcgaaggaatctccacccctgga	907

27.5%; Score 813; DB 4; Length 7333;

[illegible]

```

RESULT 4
US-09-376-781-8
: Sequence 8, Application US/09376781
: Patent No. 6261806
: GENERAL INFORMATION:
: APPLICANT: Banerjee, Papia T.
: APPLICANT: Banerjee, Clive
: APPLICANT: Andersson, Goran K.
: TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
: Patent No. 6261806
: TITLE OF INVENTION: Use
: FILE REFERENCE: 61750-267
: CURRENT APPLICATION NUMBER: US/09/376,781
: CURRENT FILING DATE: 1999-08-18
: EARLIER APPLICATION NUMBER: 60/097,015
: EARLIER FILING DATE: 1998-08-18
: NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0

```

```

RESULT 5
US-09-376-781-10
; Sequence 10, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 278
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: sequence of patl. of FOEV env gene corresponding to
; OTHER INFORMATION: residues 6057-6333 of WO 97/40167 for comparison.
; US-09-376-781-10

```

```

Query Match: 4.6%; Score 137; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 5, le-52;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 801 gccatctctccagaccggtaaattctctgttaattccggccggcaagta 860
|||||
Db 1 gccatctctccagaccggtaaattctctgttcaatccggccggcaagta 60
|||||

QY 861 caaatgatgaactataaagaataagctgtgtcccatcgacttagattcttaa 920
|||||
Db 61 caaatgatgaactataaagaataagctgtgtcccatcgacttagattcttaa 120
|||||

```


OY 921 gataagttcactgaa 937
|||||
Db 121 gataagttcactgaa 137

RESULT 6

US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 3.3%; Score 98; DB 4; Length 8132;
Best Local Similarity 100.0%; Pred. No. 1.1e-34;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2846 caccggggccgagctcttaccctcgcgtgtgtgaagactgtggcccgagcgctt 2905
|||||
Db 8035 CACTCGGGGGCGGAGCTCTTACCCCTGCTGTACGACGTGGGCCCGCCGCGCGCTT 8094

OY 2906 ggaataaaaaatcctctgtctgttgcatacaaaaaa 2943
|||||
Db 8095 GGAATAAAAAATCCTCTGTCTTGCATCAAAAAA 8132

RESULT 7

US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 3.1%; Score 91; DB 4; Length 8060;
Best Local Similarity 100.0%; Pred. No. 1.5e-31;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2846 cactcggggccgagctcttaccctcgcgtgtgtgaagactgtggcccgagcgctt 2905
|||||
Db 2501 CACTCGGGGGCGGAGCTCTTACCCCTGCTGTACGACGTGGGCCCGCCGCGCGCTT 2560

OY 2906 ggaataaaaaatcctctgtctgttgcataca 2936
|||||
Db 2561 GGAATAAAAAATCCTCTGTCTTGCATCAA 2591

RESULT 8

US-09-376-781-24
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; EARLIER APPLICATION NUMBER: 1999-08-18
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
; OTHER INFORMATION: Sequence of the 3' end of the pRV-D env region.
US-09-376-781-24

Query Match 0.9%; Score 28; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1135 gccatcagagactccatgaacaagctta 1162

RESULT 9

US-09-360-197-13
; Sequence 13, Application US/09360197
; Patent No. 6287859
; GENERAL INFORMATION:
; APPLICANT: Bassilana, Frederic
; APPLICANT: Lazdunski, Michel
; APPLICANT: Waldmann, Rainer
; APPLICANT: Deweille, Jan R.
; TITLE OF INVENTION: Human and Rat Families of Neuronal Acid-Sensitive
; FILE REFERENCE: 989.6706P
; CURRENT APPLICATION NUMBER: US/09/360,197
; CURRENT FILING DATE: 1997-07-23
; PRIOR APPLICATION NUMBER: 09/129,758
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: 60/095,408
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-360-197-13

Query Match 0.9%; Score 28; DB 4; Length 1736;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2929 tgcatcaaaaaaaaaaaaaaaaaa 2956
|||||
Db 1709 tgcatcaaaaaaaaaaaaaaaaaa 1736

RESULT 10

US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; FILE REFERENCE: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 11

US-09-376-781-30
; Sequence 30, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; FILE REFERENCE: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: consensus sequence of PERV-D.
US-09-376-781-30

Query Match 0.9%; Score 28; DB 4; Length 2000;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2047 gccatcagagactccatgaacaagctta 2074
|||||
Db 1654 gccatcagagactccatgaacaagctta 1681

RESULT 12

US-09-108-020-11
; Sequence 11, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 11

LENGTH: 1587
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-108-020-11

Query Match
Best Local Similarity 100.0%; Score 27; DB 3; Length 1587;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2930 gcatcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 1553 gcatcaaaaaaaaaaaaaaaaaa 1579

RESULT 13
US-08-831-570-2/c
Sequence 2, Application US/08831570
Patent No. 5959175

GENERAL INFORMATION:
APPLICANT: Thomas, Terry L.
APPLICANT: Nuber, Andrew N.
APPLICANT: Beremand, Phillip D.
TITLE OF INVENTION: A SUNFLOWER ALBUMIN 5' REGULATORY REGION
TITLE OF INVENTION: FOR THE MODIFICATION OF PLANT SEED LIPID
NUMBER OF INVENTION: COMPOSITION
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,570
FILING DATE: 09-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DiGioglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10545
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 743-4366
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-831-570-2

Query Match
Best Local Similarity 100.0%; Score 26; DB 2; Length 843;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2931 catcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 305 catcaaaaaaaaaaaaaaaaaa 280

RESULT 14
US-09-248-335-35
Sequence 35, Application US/09248335
Patent No. 6096504

GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEF, DANIEL
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1128-A
CURRENT APPLICATION NUMBER: US/09/248,335
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 08/924,759
EARLIER FILING DATE: 1997-September-05
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 35
LENGTH: 1098
TYPE: DNA
ORGANISM: maize
US-09-248-335-35

Query Match
Best Local Similarity 100.0%; Score 26; DB 3; Length 1098;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2931 catcaaaaaaaaaaaaaaaaaa 2956
|||||
DB 1006 catcaaaaaaaaaaaaaaaaaa 1031

RESULT 15
US-08-725-532A-2/c
Sequence 2, Application US/08725532A
Patent No. 6020179
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,532A
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0130 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-725-532A-2

Query Match 0.9%; Score 26; DB 3; Length 1275;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2931 catcaaaaaaaaaaaaaaaaaa 2956
 ||||||||||||||||||
 Db 40 CATCAAAAAAAAAAAAAAAAAA 15

Search completed: February 24, 2002, 07:31:17
 Job time: 16486 sec

QY 22 gatgcctctcttgcctgagatacaccctagccaacccttttgccttgcgaatg 81

Db 2970 GATGCTCTCTGCTGAGATTACACCCACTAGCCAAACACTTTTTCCTTCGANTGG 3029
QY 82 agagatccagtgtaacggaaacacggcagctcaactgacccgactgcccagaaggttc 141
Db 3030 AGAGATCCAGGTACGGGAAGAACCGGAGCTACCTGAGCCGACTGGCCCAAGGGTTC 3089
QY 142 aagaatcccccagccatcttgaacgaagccctacacagagactggccaattcagatc 201
Db 3090 AAGAATCTCCCGACCACTTTTGAAGAACCCCTACACAGAGACTGGCCCACTTCAGATC 3149
QY 202 caaacaccccaagtgacccctccactgactgtaagctgactgctctgaggaagccacc 261
Db 3150 CAACACCTCAGTGAACCTCTCTCCAGTACGTGATGACCTCTCTGCGGGAGCCACC 3209
QY 262 aaacagagactgcttagaaggtacgaagacactgctgtaattgctgaacctagctac 321
Db 3210 AAACAGAGACTGCTTAGAAGGACGACGAAGGCACTAGCTGGATTTCTGACCTTAGCTAC 3269
QY 322 aagaactctgttaagaagggccagatttgcaaggaaggtatacaacttgggtacagt 381
Db 3270 AAGGCTCTGTCTAAGAAAGGCCAGATTGTCAGAGAGAGAGTAACTTTGGCGTACGT 3329
QY 382 ttgcggggcggcagcagatgctgacggagagcaggaagaaactgtatccagatcag 441
Db 3330 TTGCGGGAGCGGGCAGAGATGGCTGACGAGGACGGAAGAAACTGTAGTCCAGATACCG 3389
QY 442 gcccaacacacagcaacaaagtgagagagttttgggagacgtgtgaatttgacagt 501
Db 3390 GCCCAACACCAAGCCAAACAAATGAGAGATTGTTGGGAGACAGTGGATTGTCAGACTG 3449
QY 502 tggatcccgagggtttgagaccttagacgcccactctacccgcttaacaaagaaagg 561
Db 3450 TGGATCCCGGGGTTTGGCACTTACAGACCCCACTACCCGCTAACCAAGAAAGAGG 3509
QY 562 ggaattccctgggctcctgagacacagaaaggaattgattgctatcaaaaggccctgtg 621
Db 3510 GAATTCCTCTGGGCTCTGAGACACAGAAAGGATTGATGCTATCAAAAGGCCCCCTGCTG 3569
QY 622 agcgactgctctgagccctccctgagcgttaactaaaccttacccttattgtgagt 681
Db 3570 AGCGACCTGCTCTGGCCCTCCCTGACCTAACTAAACCCCTTATGCTGATGAGAG 3629
QY 682 cgtagaggagtagcccgaggagtttaacccaacccctagagacataggaagagccgtt 741
Db 3630 CGTAGAGGAGTAGCCCGGGGAGTTTAAACCAAAACCTAGAGACATGAGAGAACCTGTC 3689
QY 742 gactacctgtcaaaagaactgtatcctgttagccagtgtgtgagccgtatgtctaaagct 801
Db 3690 GCTACCTGTCAAAAGAACTGATCCTGTAGCCAGTGGTGGCCATATGCTGTAAGGCT 3749
QY 802 atcgagactgtggccatactatgttcaagaagcgtcaaaattgactttgggagagaatata 861
Db 3750 ATCGAGCTGTGGCCATCTGCTCAAGAGACCTGACAAATTTGACTTTGGGACAAATATA 3809
QY 862 actgtaataagcccccatgcatgtgagaacatgctgcgcaagcccccaagacgaatg 921
Db 3810 ACTGTAATAGCCCCCATGATGTGAGAACATCTTGGGCAAGCCCCCAAGACGATGATG 3869
QY 922 accaagcccgcatgacccactatcaaaagcctgtcttcaagaagaggtcaacttgcct 981
Db 3870 ACCAAGCCCGCATGACCCACTATCAAAAGCTGCTTCTCACAGAGAGAGGTACAGTTCCT 3929
QY 982 ccacacagcgctcaaacctcgtgcaactcttctgctgtaagaagactgtatgaacagtgact 1041
Db 3930 CCACACAGCGCTCTCAACCCCTGCACTCTTCTGCTGAAGATGTGATGAACACAGTACT 3989
QY 1042 catgattgccatcaactatgattgagagagactgggttcgcaagagcccttaagacata 1101
Db 3990 CATGATTGCCATCACTATTGATTGAGAGAGACTGGGTCCGCAAGGACCTTTACAGACATA 4049
QY 1102 ccgctgactggaagaagtgtcaacctgttcaactgacgaagacagctatgtgtggaaggt 1161
Db 4050 CCGCTGACTGGAGAAAGTCTAACCTGTTCACTAGCAGCAAGCACTATGTGTGGAAGGT 4109

QY 1162 aagaagatgctggggcggcagtggtgagacgggagcccccaagatctgggccaagcgtg 1221
Db 4110 AAGAGAGATGCTGGGGCGCGCGGTGTGACCGGACCCGACGATCTGGCCAGAGCTTC 4169
QY 1222 ccggaaggaactcaagcgcaaaagcgtgagctcaatggccctcaacgaagcttggcgctg 1281
Db 4170 CCGGAAGGAACCTTAGCACAAGAGCTGAGTCTATGGCCCTCACAGCAAGCTTTCGGCTG 4229
QY 1282 gccgaagggaaatccataaactatatacggagacagcagatagccttgcgactgcaac 1341
Db 4230 GCCGAAGGGAATCCATTAACATTTATACGACACGAGTATGCTTTCGACTGACAC 4289
QY 1342 gtacaagggccactataacaaaggggttgcttaccctcaacgaaggaagaaataag 1401
Db 4290 GTACATGGGGCATCTTAAACAAAGGGGGTGTGCTTACCTACGAGGAGGAAATTAAG 4349
QY 1402 aacaaagagaaatctcaagcctatagaaagccttaccattgccaataaagcctagctat 1461
Db 4350 AACAAAGAGAAATTTTAAGCCTATTAGAACCGGTACATTACCAAAAAGCGTAGCTATT 4409
QY 1462 atacactgtctgactatcaagaaagcccaagatcctatctagagggaaacagatgct 1521
Db 4410 ATACACTGTCTTGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGGCT 4469
QY 1522 gaccgggttgccaagcagcagcccaagcgtcttacccttgcctataatagaagagccc 1581
Db 4470 GACCGGGTTGCCAAGCAGCGGAGCCAGGCTGTAACTTGTGCTTATATAGAAAGCCC 4529
QY 1582 aaagccccaagaaacccagacagatcacaccttagaagactgycaaagagataaaaaagata 1641
Db 4530 AAAGCCCAAGAACCCACACAGCAGTACACCTTGAAGAGACTGCAAGAGATTAATAAAGATA 4589
QY 1642 gaccagttctctgagactccggaggagacctgtatcctcatatagggaaagaaacctg 1701
Db 4590 GACCAAGTCTCTGAGACTCCGGAAGGACTGCTATACCTCAATGGAAGAAATCCGTG 4649
QY 1702 ccccaacaaagaaaggttagaatatgtccaacaaatcatcgttcaacccaacctaagact 1761
Db 4650 CCCCAACAAAGAAAGGTTAGAAATGTGTCCAAACATATCATGCTTAACCCACCTAGAACT 4709
QY 1762 aaacactgacgaagtgtgtcagaacatcccttcatcattgttctgaggtaccagagtg 1821
Db 4710 AAACACTGTGACGAGATTTGGTCAAGACATCCCTTATCATGTTCTGAGCTACAGAGAGTG 4769
QY 1822 gctgactcgtgtgtcaaacatgtgtgacctgacactgtgtaagtctatccctcaga 1881
Db 4770 GCTGACTCGGTGTCAAACTTTGTGTGCTTGCAGCTGTGTAATGCTAATCCTTCCAGA 4829
QY 1882 atacctcagaagaagactaaagggaaagccacccagcgctcaactggaagtggacttc 1941
Db 4830 ATGCTTCCAGGAAAGAGACTAAAGGGAGACCAACCCAGGCGTCACTGGGAATGTGACTTC 4889
QY 1942 actgaagtaagccgcttaatacggaaacaaatatctatgtgttctgttagaacctt 2001
Db 4890 ACTGAGGTAAAGCGCGCTTAATAACGGAACCAATATCTATTGTTTGTATACACCTTT 4949
QY 2002 tcaagatggtagagctatccctactaaagaagagacttcaacggtggtagaagaa 2061
Db 4950 TCAAGATGGGTAGAGCTTATCTTAAAGAAAGAGACTTCAACGCTGTGTGCTAAAGAA 5009
QY 2062 atactgaggaatatttccaaagatttggaaactaaaggtlaataggttcaagacaatggt 2121
Db 5010 ATACTGAGAAATTTTTTCAAGATTGTGAATACCTAAGGTAAATGAGGTACAGACATGT 5069
QY 2122 ccaagcttctgttgcacaggttaagtcagagactgccaagatatttggagattgttgaaa 2181
Db 5070 CCAAGCTTTGTTGCCAGGTATGTCAGGAGCTGCGCAAGATATTGGGATTTGATGAAA 5129
QY 2182 ctgcaattgtgatacagaccccaaggtccagagacggttagagagagatagaatacaact 2241
Db 5130 CTGCAATTGTGATACGAGACCCCAAGCTCAGAGACAGGTAGAGAGATGAATGAACCAATT 5189

Query Match	72.9%	Score 2420.6;	DB 4;	Length 8060;
Best Local Similarity	97.5%	Pred. No. 0;		
Matches 2502;	Conservative	0;	Mismatches 59;	Indels 6; Gaps
<p>GENERAL INFORMATION:</p> <p>APPLICANT: Jay A. Fishman</p> <p>TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS</p> <p>TITLE OF INVENTION: AND METHODS OF USE</p> <p>NUMBER OF SEQUENCES: 74</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: LAHIVE & COCKFIELD, LLP</p> <p>STREET: 60 State Street</p> <p>CITY: Boston</p> <p>STATE: Massachusetts</p> <p>COUNTRY: USA</p> <p>ZIP: 02109-1875</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: PatentIn Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/766,528</p> <p>FILING DATE:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/572,645</p> <p>FILING DATE: 14-DEC-1995</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Louis Myers</p> <p>REGISTRATION NUMBER: 35,965</p> <p>REFERENCE/DOCKET NUMBER: MCP-038CP</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (617)227-7400</p> <p>TELEFAX: (617)227-5941</p> <p>INFORMATION FOR SEQ ID NO: 1:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 8060 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: Single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: gDNA</p> <p>US-08-766-528-1</p>				
QY 22	gatgcctctctctgcgtgagatcaacccacatagccacacacttttgccttcgaatg	81		
DB 5495	gATGCCTTCTTCTGCGTGAATTAACCCCACTAGCCACCACTTTTGCTTGGAATG	5554		
QY 82	agagatccagatgcggaagaacccgagcagctccactgagccagacttgcctccaaaggttc	141		
DB 5555	AGaATCCAGATGAGGAGAACCCGGGAGCTCACCCTGGACCCGACAGCTCCCAAGGGTTC	5614		
QY 142	aagaactcccgacacalccttgacgaagccctacaacagagaccttgcgcaactttagatc	201		
DB 5615	AAGAaCTCCCGACaCTTGTGACGAAGCCCTTAACAGAGaCTtGGCCAACTTAAGATC	5674		
QY 202	caacacctgaagtgaccctccctcagaiacttgatgatcgtctctcttgcgaggaacc	261		
DB 5675	CAaCCCTCaGGTaGACCTCTCTCaGTAaGTGaGATGaGACCTTCTGCGGAGCaCC	5734		
QY 262	aaacagagactcttagaagtlacgaaggaacactactgccttgaaatttctgcgacttgatc	321		
DB 5735	AAaCaGAGaCTCTTAAGAAAGCaGAGCaAGCaCTaCTCTGGAATtGTGaCCTAGaGCTAC	5794		
QY 322	aaagctcttgtaagaagccagatttgcagagagagtlacacttgggttgcagtc	381		
DB 5795	AAAGCCTCTGCTaGAAGAGCCCaGATTTGaGAGAGAGAGTaACTTGGGGTaCAGT	5854		
QY 382	ttgcgagggcgagcgagatgcttgcagagagcaggaagaaacgttgatccagttaccg	441		
DB 5855	TTaCGGAGaCGGCaCGATGCTGaCGAGAGCaGGAaGAaAACTGTaTCTCaGATaCCG	5914		
QY 442	gccccaaccacagccaacaaagtgagagagatttggggagacagcttgatattgcagactg	501		

Db	5915	CCCCCAACACACAGCCAAACAAATGAGAGATTTTGGGACAGACGTGATTTTGGACAGCTG	5974
Qy	502	tgaatcccgaggttttgacacttagcagcccacacttaccgcgtacacaaagaanaagg	561
Db	5975	TGGATCCCGGGGTTTGGACCTTAGACAGCCCACTTACCCTTAACCAAAACAAAAGG	6034
Qy	562	ggaattcccccgggtctctgacgaacgaagcaattgactatcaaaaagccctgtg	621
Db	6035	GAATTCTCTGGGCTCCTTAGACACCGAAGGAGATTGATGATCAAAAAGGCTGTG	6094
Qy	622	agcgacactgtcttggccctccctgacgtaacatacccttacccttatgtagatag	681
Db	6095	AGCGACACTGCTTGCCCTCCCTGACGTAACCTTATACCTTTATGTGATGAG	6154
Qy	682	cgtaaggagtagcccgagagatttaaacccaaacccctaggacataggaagagccgtt	741
Db	6155	CGTAAGGAGATGACCCCGGGGAGTTTAAACCCAAACCTTAGACCATGAGACAGCTGTC	6214
Qy	742	ggcttaccttcaagaagcttgatctctttagccagttgttgcccgtatgtctgaagct	801
Db	6215	GCCTACTCTCAAAAGACTGATCTGTATCCACAGTGGTGGCCATATGCTGAAGGCT	6274
Qy	802	atccgacgtctgtagcatactggtcaagagcctaacatitgactttggagc- agaatat	860
Db	6275	ATCGCAGCTGTGGCCATCTGATGCAAGACGCTGACAAATTGACTTTGGACAAAGAAATT	6334
Qy	861	aactttaagcccccaatgacttgatggagaaatcgtttcggcagcccccaacgaatgtagt	920
Db	6335	AACGTATATAGCCCCCATGCAATGATTTGGAACATCGTTGGCAGACCCCGACAGATGAT	6394
Qy	921	gaccaagcccgcatgacccaactatacaagccgctcttcaagagagaggtaactttgcg	980
Db	6395	GACCAACGCCCGCATGACCCACATATCAAAACCTGCTTTCACAGAAAGGCTACGTTGGC	6454
Qy	981	tcacacagccgctctaaccttcccaactcttgcctgaagagactgtaacaaagttagc	1040
Db	6455	TTCACCAACCGCTTCAACCTGCACTTCTGCTGTAAGAGACTGATGAAACCACTGATAC	6514
Qy	1041	tcatgattgcataactatattgattgaagagacttgggttcgcaagagacctaacagat	1100
Db	6515	TCATGATTGGCATCAACTATTGATTGATGAGAGACTGGGGTCCGCAAGACCTTTAAGACAT	6574
Qy	1101	accgttgcacgtgaggaagtgttaactggttcaactgacggaagcaagtgtgtgtgaagg	1160
Db	6575	ACCGCTGACTGGAGAAGTGTCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGAAGG	6634
Qy	1161	taagaggaatgctcgtggcgcgagttgttgacogggagcccgcaagatcttgggcacagcct	1220
Db	6635	TAAAGAGATGGCTGGGGCGGCGGTGTGAGCGGACCCGACGATTCGGGCCACAGCCT	6694
Qy	1221	ggcgggaaggaactcaagcgcaaaagcctgaagctcatgcccctacgcaagcctttgcgct	1280
Db	6695	GCGGGAGAGAACTTCAGACAAAAGGCTATGGCCCTTACGCGAACACTTTGGGGCT	6754
Qy	1281	ggcggaaggaaatccataaacaattatagagacagaagtagtgcctttgacgtcacaa	1340
Db	6755	GGCCGAAGGGAATCATTAACATTATTATCGGACAGCAGGATATGCCCTTTCGACTGCACA	6814
Qy	1341	gctacacggggtcactataacaagaagggttctaactcaagggaggaagaaataaa	1400
Db	6815	CTCATATGGGCGCATCTTTAAACAAAGGGGCTTCACTCAACGAGGAGGGAATAATAA	6874
Qy	1401	gaacaaagaagaaattctaagcctatagaagccttaccttccaaaaagcctagat	1460
Db	6875	GAACAAGAGGAATTTTAAGCCATTATAGAGCGTAACTTTAACAAAAGGCTATGCTAT	6934
Qy	1461	tataacactgtctcgtgacatcagaagaacccaagcttcacatcattagaagggaacccaatg	1520
Db	6935	TATACACTGCTCTGACATATAGAAAGCTTAAAGTTCATATCAAGAGAAACAGATGGC	6994
Qy	1521	tgaccgggttgcacagcagcagccagcgtgttaaccttgcctataatagaagcc	1580

Db	6995	TCGACGGGTTC	CAAGCAGCAGCCAGCCAGGGGTGTTAACTTTGCGCTATATAGAAATGCC	7054
QY	1581	caagccccaga	accaccagacagtaacaacctagagaagctggcaagaagataaagaat	1640
Db	7055	CAAAAGCCCCA	AAACCCAGACCCAGACGACGATACACCTTAGAAGACTGGCAAGAGATAAAAAAT	7114
QY	1641	agaccagttc	ctcctcagagctccgaagggacctcgtataaccataatgggaagaatcc	1700
Db	7115	AGACCA	-TTCTCTGAGACTCCGGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCCT	7173
QY	1701	gccccacaaga	agggtttagaatalatgccaac-agaatacatcgtctaaacccactagaa	1759
Db	7174	GCCCCAACAA	GAGGGGTTAGAAATATGTCCAAACAAATACATCCTCAACCCACCTTAGAA	7233
QY	1760	ctaaacccctc	gcagagcttggfccaagaacatccccctatcatgltctcctcagag	1819
Db	7234	CTTAACACCT	CGCAGCATGGTGTCACAAACATCCCTTATCATGTTCTGAGGCTACAGAG	7293
QY	1820	tggctgactc	ggtgtccaacatctgtgacctccagcctcgtgtaattgcttaacctca	1879
Db	7294	TGGCTGAC	TGGGTGCAAACTTTGTGTGCTCTGCGCTGGCTGTTAAATCTTTCA	7353
QY	1880	gaatacctcc	aggaagaagactaaagggaagccaccagcgcctcaacttgggaagt	1939
Db	7354	GAAATGCTCT	CCAGGGAAGAGACTTAAGGGGAACCCACCAGCGGCTCACTGGGAATGGACT	7413
QY	1940	tcaatpagg	laagccgcgtcaataacggaacaaatctcatggtttctgtagacact	1999
Db	7414	TCACTGAG	GTAAAGCCGGCTAAATATGTGAACAAATACCTATTTGGTTTGTAGACACT	7473
QY	2000	tttcagat	tggttagaggtctatccctaagaagaagacttcaaccgttggcttaga	2059
Db	7474	TTTCAGAT	GTGGTGAAGGCTTATGAGGCTTATCTCAAGAAAGAGACTTCAACCGTGTAGCTAAA	7533
QY	2060	aaatactg	aggaaaattttcccaagatttgaatcttaagtaagtaaggtcagaacatg	2119
Db	7534	AAATCACT	GTGAGAAATTTTCCAAAGATTGTGAATACCTAAGGTAATATAGGTCACAGATG	7593
QY	2120	gtccaagct	tccgttgcccaagtagtaagtaaggaacttggccaagatatttgggaatttaga	2179
Db	7594	GTCCAGCT	TTTGTGTCGCCAGGTAAAGCAGGGAGCTGGCAAGATTTGGGGATTGATTGGGA	7653
QY	2180	aactgcatt	gtgcaataagaccccccaagctcagagacaggttagaagagaatgaataagcca	2239
Db	7654	AACGCAAT	TTGTGCATACAGACCCCAAAAGCTCAGGACAGGTAGAGAGATGAATGAACCA	7713
QY	2240	ttaaagag	acccttaccaatltgacacagagaaactgagatlaagtattgattgctctcc	2299
Db	7714	TTTAAAGAC	CCCTTACTAATTTGACCCGGAGACTGGCTTAAATGATTGGATACCTCTCC	7773
QY	2300	tgcacctt	gtgctttttaggttgaagaaaccccttgcagatttgggtctaacccccctatg	2359
Db	7774	TGCCCCTT	GTGTGTTTTTAGGGTTAGGAACACCCCTGACAGATTGGGCTACCCCCCTATG	7833
QY	2360	aattgtctc	agggagcccccccgcttggagaagaattgctcttgcacalagtgcfatg	2419
Db	7834	AATTACT	CTAGGGGAGACCCCCCATTGGTAGAAATTTGCTTCTGACATAGTGTGATG	7893
QY	2420	tgtcgtct	ccagccttltgtctctatgctcgaagtgctcgaagtgggttgggaagcagag	2479
Db	7894	TGCTGCTT	TCCAGGCTTTGTTCTCTAGGCTCAAGGCACTTTGATGGGTGAGACAAAG	7953
QY	2480	cgltgaag	cagctcccggaaggtcctacac---aggaagaaacttgcagtctccacatcgct	2536
Db	7954	CGTGAAGCA	CTCCGGAGGCTCACTCAAGAGAGAGAACTTGCAGATCCACATGCTT	8013
QY	2537	tccaagt	tggaattcagttcagttatgttgaagcagccgltcagaagaa	2583
Db	8014	TCCAAGTG	GAGATTCAAGTCAAGTTAGAGAGCCACCGTTCAGGAAC	8060

RESULT 3
US-08-766-528-2

Sequence 2, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7333 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 53.4%; Score 1771.8; DB 4; Length 7333;
Best local similarity 95.1%; Pred. No. 0; Mismatches 17; Indels 79; Gaps 1;
Matches 1871; Conservative 0;

QY 1351 gcatctataaacaaggggtgtgtcttaccagcagggaggaataaagaacaagag 1410
DB 3504 GCTAACCAAGAAAAAGGGGTTGCTTACCTCAGCAGGGAGGGAATTAAGCAAAAGAG 3563
QY 1411 gaaattcaagcctattagaagccttaactttgccaanaaggctagctattatacaact 1470
DB 3564 GAAATCTTAAGCCTATTAGAGGCTTACATTGGCCAAAAGGCTAGCTATTATACAGT 3623
QY 1471 ccttgagatagaagaaccaaagatccatatactagaagggaacagaatgctgacgggtt 1530
DB 3624 CCTGGAATCAACGAAGCAAGCAAGATCTATATAGAGGAGACAGATGCTGCTAAGCGGTT 3683
QY 1531 gccaaagcagcagccagcgtgttaacctctgctcataatagaacgcgccaaagccca 1590
DB 3684 GCCAAGCGAGCGACGCCAGGGCTGTATACCTCTGTGCTATATATAGAAACGCCCAAGCCCA 3743
QY 1591 gaaccagagacagatcaccccttagaagactgagcaagagataaaaaagatagacagttc 1650
DB 3744 GAACCAAGACAGACTACACCTTAGAAGACTGGCAAGAGATTAATAAAGATAGACCGTTTC 3803
QY 1651 tctgagactccggaggagactgtctatacctcatataggaagaaatccctgcccaca 1710
DB 3804 TCTGAGACTCCGGAGGAGACCTGCTATATCTCATATGGGAAGAAATCTGCCCCACAAA 3863
QY 1711 gaagggttagaatgtccaacagatacatcgtctaaaccaactaggaactaaacaactg 1770
DB 3864 GAAGGTTAGATATGTCTCCAAACAGATACATGCTTAACCCACTAGGAATTAACACCTG 3923

QY 1771 cagcagttgtgcagaacatcccccttaacatgcttctgaagcttaccagagtgctgactcg 1830
DB 3924 CAGCAGTTGTGTCAGAAACATCCCTTATCTATGAGGTACTACAGAGAGTGCTGACTCG 3983
QY 1831 gtgttcaaacatgtgtgtccctgacagctgtgtatgttaacttccccaataactcca 1890
DB 3984 GTGTCAAAACATGTGTGTGCTCCCTGCGAGCTGTATGTATGCTTACCTTCAGAAATCTCCA 4043
QY 1891 ggaagaagactaaggagggaacccacagcgctcactctgaggagtgagcttaactgaqta 1950
DB 4044 GGAAGAAGACTAAGGAGGAGCCACCCAGCGCTCATCTGGAAAGTGAGCTTACTAGAGTA 4103
QY 1951 aagcggctcaataacaggaacaaataatctatgtgttcttgaagacatttcagatgg 2010
DB 4104 AAGCGGCTAAATACGGAACAAATATCTATTGGTTTTGTGACACCTTTTCAGAGATGG 4163
QY 2011 gtagagcttatctactaagaagaagactcaacggtgtgtgtgtctaaagaataactgag 2070
DB 4164 GTAGAGGCTTATCTCTATAAAAAAGAGACTTCAACCGTGTGTGCTAAGAAAATCTGGAG 4223
QY 2071 gaaattttccaaagatttgaataacctaaagtaaggtcagacaatgtgtccagcttc 2130
DB 4224 GAAATTTTCCAAAGATTGGAATTACTAAGTAAATAGGGTCAGACAAATGTGCACCTTTC 4283
QY 2131 gttgccaggtlaagtcaggagactggtccaagatatgtgggattgattggaactgtcattgt 2190
DB 4284 GTTGCCAGGTAAGTCAAGGACTGGCCAAAGATTTGGGGATTGATTGAATACTGCATTGT 4343
QY 2191 gcatacagaccccaagctcagagcaggttagaagagatgaatgaacccatlaaagacc 2250
DB 4344 GCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATGAACCAATTAAGAAGACC 4403
QY 2251 cttaaccaattgacacacagagactgtgcatlaatgtatgtgtgtctctccgtctgtg 2310
DB 4404 CTTACCAATTGACACACAGAGACTGTGCAATTAATGATGGATGCTCCTCCCTTTGTG 4463
QY 2311 ctttttaggtgaggaacaccccttgagacgttttggtctgaaccccttaagaattgtctac 2370
DB 4464 CTTTTAGGGTGTAGAGAACCCCTTGACAGATTTGGGGTGAGCCCTTAATAATGCTCTAC 4523
QY 2371 gggggacccccccctgtgacagaatgtgctttgcacatagtgctgtgtgtctgtcttc 2430
DB 4524 GGGGGACCCCCCCTGTGACAGAAATTCCTTTGACATAGCTGTGATGTCTCTCTTTC 4583
QY 2431 cagccttgttctctagagctcaagcgctcaggtgtgtgtgagcagcgatgaagcag 2490
DB 4584 CAGCCTTGTCTCTAGAGCTCAAGCGCTCAGTGGGTGAGGACAGCAGCGTGGAAAGCAG 4643
QY 2491 ctcgggagagcctactcagagagagactgtgaagtccacatcgcttccaagttagat 2550
DB 4644 CTCGGGAGGCGCTACTCAGAGAGAGACTGTCAAGTTCCATCGCTTCGMAATTGAGAT 4703
QY 2551 tcagctatgtttgaagcgcccggtgcaagaaacctcgaactcgtgtgaaaggacctat 2610
DB 4704 TCACTATATGTGTGAGCGCCACCGTGCAGGAACCTCTGAGACTCTGTAGAAAGGACCTTAT 4763
QY 2611 ctgctacttlttgacacacaaagcggtgtgaaagtctgaaggaatcccaaccgtgtccat 2670
DB 4764 CTGCTACTTTTGACACACACACACGCGGTGTAAAGTGTGAAG----- 4804
QY 2671 gcatccacgtttaagccggcgccacctccgattcgggtgtgaaagcgaaagactgaa 2730
DB 4805 ----- 4804
QY 2731 aatcccttaagctcgtccctccacgctgtgttccctactctgttaataactctcaagt 2790
DB 4805 AATCCCTTAAGCTTCCGCTCCATCGCTGTGTCTTATCTGTGCAATTAACCTCTCAAGT 4864
QY 2791 taatgttaagccttftggaacagcccgaaactccataaacccttatctcactcgtgt 2850
DB 4865 TAATGTAAACGCTTGTGTGACAGCGCAACTCTCCATTAACCCCTTATCTCTCACTGTGTT 4924

```

QY 2851 aactagactccggtacaggtatattatataacagcactcaaggaggagcctcccttgg 2910
Db 4925 ACTTACTGACTCCGGTACAGGATTAATTAATTAACAGCACTCAAGAGAGGAGCCTCCCTGGG 4984
QY 2911 gacccgtggtcctgataatattatgtctgccttcgaacgaacccctgtctcaatgca 2970
Db 4985 GACCTGGTGGCTGATTAATTAATGTCTGCTTCGATGATATCCCTGTCTCATATACCA 5044
QY 2971 ggcacacaccccgatgactcgtgcttccggtttcagtttgcacagaccccaaa 3030
Db 5045 GGGCCACACCCCGGATGATCCTGCTTACGGGTTTACGTTTGGCCAGAGACCCCAAA 5104
QY 3031 taatgaagaatattgtggaatcctcagaattcccttgaagcaatgagctgataac 3090
Db 5105 TAAATGAAGAATATTGTGGAATCTCAGATTTCTTTGCAAGCAATGAGAGCTGATAC 5164
QY 3091 ttctaataatgaggaatggaatgagcagctctcagaagaacagatgatactctt 3150
Db 5165 TTCTAATGATGGAATTTGGAATGGCCAGCTCTCTCAGCAAGACAGATAAGTTACTCTT 5224
QY 3151 ttttaacatcctcactacagttataatcaatlaattatgagcagtggaatggaatg 3210
Db 5225 TGTTAACAATCCTACACAGTTATATCAATTTATATGCGCATGGAGATGGAAGATTG 5284
QY 3211 gcaacagcgggtacaaaagatgtagcaataaagaacaaatgactgactgtagagct 3270
Db 5285 GCAACAGCGGATCAAAAAGATGATGAGAAATTAAGCAATTAAGCTGATCTGTTAGACT 5344
QY 3271 agattacttaaaaatgaatttacttaaaaaaagagagagagagagagagagagag 3317
Db 5345 ACATTACTTAAATAATAGTTTCACTGAAAGAAAGAAACAAAGAAATA 5391

```

RESULT 4

```

US-09-075-272-1
Sequence 1, Application US/09075272
Patent No. 6136598
GENERAL INFORMATION:
APPLICANT: MILLER, A. DUSTY
APPLICANT: WOLGAMOT, GREG
APPLICANT: BONHAM, LYNN
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
TITLE OF INVENTION: PACKAGING CELL LINES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Poor, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-075-272-1

```

```

Query Match 35.4%; Score 1175.8; DB 3; Length 8655;
Best Local Similarity 66.1%; Pred. No. 0;
Matches 1778; Conservative 0; Mismatches 877; Indels 36; Gaps 4;

```

```

QY 19 acagatgctctctcctgctgctgagattacaccccaagcgaacacttttgcctgaa 78
Db 3095 AAGATGCTCTCTCTTTTGGCTGGCTGACACCCCAAGAGTCACTTGTGGCTTTGAA 3154
QY 79 tggagatcccaagtagcaggaagacccgagcgtcactcactgagcccgactgcccagaag 138
Db 3155 TGGAGAGACCCAGAGGGCGGACAGACTGCTCAATTAACTGGACTAGACTACCAAGGG 3214
QY 139 ttcaagaactcccgacacatctttagaagaagccctacagagccctggccaactcagg 198
Db 3215 TTTCAAAAATCCTCCCTACCTGTTTACGAGAGCCCTCCATCGGATTTAGCAACCTTTGCG 3274
QY 199 atccaaacacccctcaggtgagccctcctcagtagcagtagatgactgctctgagcagc 258
Db 3275 GCTCAAAAACCCCGACCTTACCTACTGCTAGCTAGATGATCTTTTAATCGCAGCGCC 3334
QY 259 accaaacagagactgttagaagtagaaggaagcaactactgctgtaattgctgacctagg 318
Db 3335 TCAAAAAGAGCTATGTCAACAGAGGAGCTGAGAGACTCTCAAGAGCTGGGGAATTTGGGG 3394
QY 319 lacagagcctctgcttaagaagggccagatttgcagagagagtagaactacttggagtag 378
Db 3395 TATGAGATTTCGCTTAAAGAGCACAATTTGTCAAACTGAGGTATCTACTTGGGCTAT 3454
QY 379 agtttcggggcgagcagatgctgagcagagagcaggaagaaacgtgtagtccagata 438
Db 3455 ACTTTGGAGAGAGTAAAGATGCTGCTGAGAGCCCGGAAACACTGCTATGATGAT 3514
QY 439 ccggcccaacacacacacacacagatgagagagtttttggagacagcgtgatttgcaga 498
Db 3515 CCACCGCCAAACACACACAGGAGTACGTAGTTTCTGGGAGCTGCTTTTGTAGA 3574
QY 499 ctgtgactccggggttgcagacttagcagcccaactcactccgactcaacaaagaaaa 558
Db 3575 CTCTGGATTCCAGGGCTTTGGAGCCTTACAGACCTCATATCTCTTGTAGTGGAGGA 3634
QY 559 ggggagttcctctggtcctcagacacagagagcatttgaatgctatcaaaagggccctg 618
Db 3635 ATTCCCTTTGAATGGAAGAGAACACCAAGAGCTTTTGAAGCTATCAATATGCTCTTA 3694
QY 619 ctgagcagactgctcctgcccctcctgagtagtaaccccttaacccttaattgtagat 678
Db 3695 ATGACTGCGCTCGCGTAGATTTACAGACTTGTCAAGCTCTTCTCATATGTGGAGC 3754
QY 679 gagcgtlaagagtagcccgagagagtttaacccaaacccctagacatgagagggagacct 738
Db 3755 GAGAGAGCGGGCATATGACCAAGGGGTGCTGACAAAGCACTAGAGACCTTGGAGAGACCG 3814
QY 739 gtgctcactgtcaaaagagagctgtgctgtagcaggtggttggccglatgttgaag 798
Db 3815 GTAGCTTATCTGTCAAAAACACTGGAATCCGTTGCTAGTGGATGGCCCAATGTCTGAAA 3874
QY 799 gctatgcagctgtgcatcactgtgcaagagcagcgtgacaatltgacttggagagaat 858
Db 3875 GCTATTGCAAGCACTAGCTCTGCTGATCAAAAGTGTGTAATCAATGCAATGGACAGAG 3934
QY 859 ataacgtlaagcccccatgcatgtgagacacatcgcttgcagagcccccagagcagtag 918
Db 3935 GTGAGTGTGCTGGCCCCCTTATGCTCTGAGAGTATTGGCGAGCCACCTGATAGATGG 3994
QY 919 atgaccaacgcccagcatgacccatcaagaagcctgcttccacagagagagtaacttc 978

```

Db 3995 ATGATAATGACCGGATGACACACTATCAGATCTGCTGCTAAATGACCGGTAACCTT 4054
 QY gttccaccagcgcctctcaaccctgcacactctctgctgctgaagactgatacgaagc 1038
 Db 4055 GCGCCCTCTGCATTTCTCAACCCAGCCACCTCTCTCTTAACGATGATTCGCTCCCA 4114
 QY 1039 actcattgctgcaatcactatctgattgagagagcctgagctccgcaagacccctacagc 1098
 Db 4115 GTACATGATGTGACAGACATCTGCGGTGAGAAATTTGGGACAGAAAGACCTGACTGAC 4174
 QY 1099 ataccgctgactgagagaagtgtcaacctgttactgacggaagcagctatgtgtgaa 1158
 Db 4175 CAACCTGCGCTG---AGCTCTAGTGTGTACACGATGCGAGTGGAGTTTCTGATTTGAG 4231
 QY 1159 gtaagagagatgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1218
 Db 4232 GGAAGAGGAGGCTGAGCTGCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4291
 QY 1219 ctgacggaaggaactcactgacgcaaaagctgagctcagcctcagcagcagcttgcg 1278
 Db 4292 TTGCTGAGAGAACTTGGCAGACAGAAAGGCTGAACCTCATAGCGCTTACACAGGCGCTCCGA 4351
 QY 1279 ctgacggaaggaatcactcaaaacttatacgaagcagagtaagccttgcgactgca 1338
 Db 4352 GAGGCGCAAGGTAAGATCAATTAATTTACACTGACAGCGCTATGCTTTGCTACTGCA 4411
 QY 1339 cagctacacgggacatcactaacaaggggtgtgcttaccctcagcagagggaggaata 1398
 Db 4412 CATATCATATGGGCGCATCTACAGGACGAGGCTTATTCAGCTTACACCGGTAAAGCATTC 4471
 QY 1399 aagaacaagaagaatctcactaagcctattagaagcctcattgtccaaaagcagctagc 1458
 Db 4472 AAAAACAAGAAAGAAATTTCTGGCGCTGTAGAGCCATACATGCGCCCAAGAGGTAAGCC 4531
 QY 1459 attatcagctgctcctgagcctgagaaagcgaagatcctatatactagaaggaagcagag 1518
 Db 4532 ATCATCTACTGTCCTCGGTCACGAAAGAGAGGAGGACTTGTGTGCGCAAGGCAACCGGATG 4591
 QY 1519 gctgacgggtgtgccaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1578
 Db 4592 GCAGACTCATGACCAAAAGCGTTGCTCAGGGGCGCATGATCTTAAGTGAAGAAAGAAAT 4651
 QY 1579 cccaagccccaagaccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1638
 Db 4652 CC-----GTCCAAAGGCCCTGA-----GGATGAAAGCATATGAT 4684
 QY 1639 atagaccagttctctgagcctcagcagcagcagcagcagcagcagcagcagcagcagc 1698
 Db 4685 ATAAAAGAACTATTTGGAGCTAGTATCCCTCCATCTTTTTCGAAAGGAAAGATATGAC 4744
 QY 1699 ctgcccacaagaagaaggttagaatactgccaacagatacatcgcttaaccacccataga 1758
 Db 4745 TTGACACCGAGAGGAAGATTAATTTGTGAAGAGACTACACAGTTTACCCACCTGGGA 4804
 QY 1759 actaaacacctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1818
 Db 4805 GTTGAAGAAATGATGAGACTGATTAAGTCCCGGATATCAAGTCCCTTAAGTGAAGTCA 4864
 QY 1819 gtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1878
 Db 4865 GTGGCTCAAAAGATTTAACTCTCTCAAAAGCGTGTGCACTTCACTAATCAACTAAACCC 4924
 QY 1879 agaatactcagaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1938
 Db 4925 TACAAAGAACTGGAAGAGACAGCGGAGAACCGTCTCGAGTGTATTTGGAGGTGAT 4984
 QY 1939 ttactgaggttaagcccgctaaatacgaagaacaaatctatgtgtttttagaacacc 1998
 Db 4985 TTACTGAGATTAACCCGGAATGTATGTAATTAAGTATCTGTATTTAGTATTTAGACACC 5044
 QY 1999 ttctcagaggttagagcttactactactactactactactactactactactactactact 2058
 Db 5045 TTTTTCAGATGGGTGTAAGCGTTCCCTCACTAAAGTGAAGATGCCAGATTGTGGCCAG 5104

QY 2059 aaatactgaggaattttccaaagatttggaatacctaagtaataagtcagacacat 2118
 Db 5105 AAGATTTTGAAGAAATCTGCAAGATATGAGTACCTAAAGTATGAGGCTCCCAAT 5164
 QY 2119 gttccagcttctggtccaggaagtaagtcaggaacggtgccaagatacttggaattatg 2178
 Db 5165 GGACCAAGCTTTGTTGCCAGGTAAAGTCAAGGATTTGGCCACTGATTTGGCATTTATTTGG 5224
 QY 2179 aaactgactgtgcatcagaccacaagcctcaggaagcagcagcagcagcagcagcagcagc 2238
 Db 5225 AAATTAATCATTTGCTTACCGCCCTCAAAAGCTCAGAGCAGGTAGAGAGATGATAGACA 5284
 QY 2239 attaaagagccttaccacaatttgaccacagagcagcagcagcagcagcagcagcagc 2298
 Db 5285 TTTAAAGGAACTTGAATTAATTTGCGCATGAGACCGCGGGAAGACTGGGTGCTTC 5344
 QY 2299 ctgacctgtgcttctttagagtgaggaacacccctgagcagcagcagcagcagcagcagc 2358
 Db 5345 CTCCCGCTTGGCTCTTCGAGACCGCGGAACACCCCGGACGTTTGGGCTCATCTCTTT 5404
 QY 2359 gaattgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2412
 Db 5405 GAAATTTCTGTATGAGGAGACCTCCCTTTAATAAAGATGAGTGAACATTGATTCGCGAT 5464
 QY 2413 gctgagtgtgcttccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2472
 Db 5465 TCAGGCTCTGTTCACCT 5524
 QY 2473 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2532
 Db 5525 ACCCAATTTGGAGACGAGCTAAAGAGCGCTACACCCAGGAGACACCGCAGTACCCAC 5584
 QY 2533 cgttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2592
 Db 5585 GAGTTCAGAGCTGGGACCAAGTCTGTGACAGCATGCAACCGGTAGCGCTTGAACCA 5644
 QY 2593 cgttgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2652
 Db 5645 CGGTGGAAGGAGACCTATTATTTAGTGTACTAACACACTCTTACGGCGAGTAAAGTGGACGG 5704
 QY 2653 atcccacacctgatacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2703
 Db 5705 ATTGCTCTGATGATCAGCTTCAACAGTCAAGAGGCGCCCGACGACGAT 5755

RESULT 5
 US-09-111-085-3
 ; Sequence 3, Application US/09111085
 ; Patent No. 610034

; GENERAL INFORMATION:
 ; APPLICANT: Stoye, Jonathan P
 ; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
 ; FILE REFERENCE: 4238/75168
 ; CURRENT APPLICATION NUMBER: US/09/111,085
 ; EARLIER APPLICATION NUMBER: GB 9710154.7
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 3482
 ; TYPE: DNA
 ; ORGANISM: Porcine retrovirus
 ; US-09-111-085-3

Query Match 34.1%; Score 1131; DB 3; Length 3482;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 1251; Conservative 0; Mismatches 130; Indels 12; Gaps 3;

OY	1767	cccaagaaagtcttgcaagaacatccctcttaactaactctcgaagcttacaagaagctgctcga	18203
Db	6	ccctgcagcagcttctgcagaacaacccctcttaactcgtctccgaagcttaccgaagctgctcga	65
OY	1827	ctcgtgtgtctcaaacatctgtctgcctctgcagctcgtgttaactgtctaactcctccaagaatacc	18866
Db	66	ctcgtgtgtctcaaacatctgtctgcctctgcagctcgtgttaactgtctaactcctccaagaatacc	125
OY	1887	ttccaggaagaagactaaaggaggagccacccagcgtctcaactctgggaagctggaattcaatcga	19466
Db	126	ttccaggaagaagactaaaggaggagccacccagcgtctcaactctgggaagctggaattcaatcga	185
OY	1947	ggttaaaagccggtctaaatacgaacgaacaatactcatctgtgttttctgaacactcttcag	20066
Db	186	ggttaaaagccggtctaaatacgaacgaacaatactcatctgtgttttctgaacactcttcag	245
OY	2007	atcggtctgaagcttatactctactcaagaagaagacttcaaccgtgtgtgtctgaagaataact	20666
Db	246	atcggtctgaagcttatactctactcaagaagaagacttcaaccgtgtgtgtctgaagaataact	305
OY	2067	ggaagaagaattttctcaagattcttgaaatactcaaggttaaggttaaggttcaagaataagctccagc	21266
Db	306	ggaagaagaattttctcaagattcttgaaatactcaaggttaaggttaaggttcaagaataagctccagc	365
OY	2127	tttcgtctgcacaggtlaagctcaagagactctgcgaagataattctggagatttgaataactctga	21866
Db	366	tttcgtctgcacaggtlaagctcaagagactctgcgaagataatttggatttggatttgaataactctga	425
OY	2187	ttctgtgcatacaagacccccaagaagctccaagaacagtgtaagagagatagaatagaacattaaaga	22466
Db	426	ttctgtgcatacaagacccccaagaagctccaagaacagtgtaagagagatagaatagaacattaaaga	485
OY	2247	gaaccttacaacaatttgaccaccaagaagactctgcatactaaatgaattctctctccctc	23066
Db	486	gaaccttacaacaatttgaccaccaagaagactctgcatactaaatgaattctctctccctc	545
OY	2307	tgctgcttttaagggtgaaggagacacccctctgacaagtttggctcgtgaccctctatgaatttgc	23666
Db	546	tgctgcttttaagggtgaaggagacacccctctgacaagtttggctcgtgaccctctatgaatttgc	605
OY	2367	ctaaagggggagacccccccgctctgcaagaatactgcctcttgacacatagtctgaatgtctgc	24266
Db	606	ctaaagggggagacccccccgctctgcaagaatagtctctctgaacatagtctgaatgtctgc	665
OY	2427	ttcccaagccttcttctctctcctcagaagcgtctgaagctgtgtgtgtgtgcagcgagcgtgtga	24866
Db	666	ttcccaagccttcttctctctcctcagaagcgtctgaagctgtgtgtgtgtgtgcagcgagcgtgtga	725
OY	2487	gcaacctccgggagagccctacacag---aagagactctgcagcttccacactgcctccaagt	25433
Db	726	gcaacctccgggagagccctacacagagaaagagagacttgcgaagttccacactgcctccaagt	785
OY	2544	ttggaagatcagctctatgttttagagcagccacccgtgcgaagaaacctcgaagactcggctggaag	26033
Db	786	ttggaagatcagctctatgttttagagcagccacccgtgcgaagaaacctcgaagactcggctggaag	845
OY	2604	accttatctcgtactcttttgacacacacacacagcgtctggaaggtcgaaggaatactccaactg	26633
Db	846	accttatctcgtactcttttgacacacacacacagcgtctggaaggtcgaaggaatactccaactg	905
OY	2664	gattccatgcatacccaagcttaagagccggcgccacactcccgattctcgggtgtggaagccggaana	27233
Db	906	gattccatgcatacccaagcttaagagctggcgccacactcccgattctcgggtgtggaagccggaana	965
OY	2724	gactcgaanaatccaccttaagctttgcctccatcagctcgtgttctcttaactctgtccaataactc	27833
Db	966	gactcgaanaatccaccttaagctttgcctccatcagctcgtgttctcttaactctgtccaataactc	1025
OY	2784	ctcgaactaaatgtgttaaacgctctgttggaagcccggaactccacataaaccttatctctca	28433
Db	1026	cccaagccagtagtaaacgcttatagacagcttcgaaccccatagaccttatctctca	1085
OY	2844	ccctggttacttactgactccggtacaggtatataataatacaagaactcaagggagagctc	29033

[illegible]

Db 3371 GATGCTTTTCTGCTCAGGCTACATCCCAACAGCCAGCGCTGTTGCTTCGATG 3430
QY 82 agagatcagtagcaggaacccgagctcaccttgaccagctgcccgaaggtc 141
Db 3431 AAGAGCCCGAAGAAAGGTACACAGCTGAGCTGAGCCGGCTACCAACAAAGGTTTC 3490
QY 142 aagaactccccacacatctttagcgaagcctacacagaagccttgacaacttcagatc 201
Db 3491 AAGAACTCTCCACTCTCTTCGACAGAGGCCCTCCACCGAGATTTGGCTCCCTTTAGGCC 3550
QY 202 caaacacctcaagtgacacctctcctcagtaagtgatgacacctgctctgagggagcaacc 261
Db 3551 CTCAACCCCGAGTGTGTACTTCAATATGTGACGACCTCTTGGCGCCGCCCA 3610
QY 262 aacaggaactgttagaagtgacgaagcactgctggaattgtcagactaggtac 321
Db 3611 TATGAAGACTGCAAAAAAGAAACACAGAAAGCTTTACAGAGATTATAGTAAGTTGGGTTAC 3670
QY 322 agagcctctgtcgaagaagcccaagatttgcaagagaagtgtaacacttgaggtagt 381
Db 3671 CGGGTATCGGCTAAGAGAGCCAGCTGCGCAGAGAGAAAGTACCTTCTGGGGTACCTA 3730
QY 382 ttgcggggcgggacagctgtgctgacgaagcaggaagaacactgtatgctcagatagc 441
Db 3731 CTCAAGGAGAGAAAGATGCTACACCCAGCCGAAAGGCTACTGTTATGAAAAATCCCT 3790
QY 442 gcccaacacacagccaacaagtgagagagattttgggagacgtgtgatttgagagctg 501
Db 3791 GTTCTACGAGCCCGCAGAGAGTCCGTGAATTTCTAGGACATGCCGGATTCTGCAAGGCTC 3850
QY 502 tggatcccgagggttggaaccttagcagcccaactcaccgctaaccaaaagaaag 561
Db 3851 TGGATCCCTGGGTTTGTCTTCCCTGCTGACCCCTTGTACCCCTTAACAAAGACAGCATC 3910
QY 562 ggaattcctctgggctccgagcaccagaagcatttgatgatalcaaaagagccctgtg 621
Db 3911 CCTTTATTTGACTGAGAGAAATACGAGGCTTTTGACCAATTAATAAAAAAGCCTTGCTG 3970
QY 622 agcgacactgtctlgccctccctgacgtaactaaaccttaccctttagttagatgag 681
Db 3971 TCAGGCCCTGCATTTGGCTCCAGAGCCTCACCAAGCATTTACTCTATATATATAGATGAG 4030
QY 682 cgtaaaggagtagcccgagaggtttaaaccaaacctagagacataggaagagacgtgt 741
Db 4031 AGAGCCGGCGTGGCCCGGGGAGTGTCTACTCAGACTTTAGAGCACTTGCGGCGCCAGTA 4090
QY 742 gcttaacctgtcaagaagctgtagtccgtgtagccagtggttggccgtagtctgaagct 801
Db 4091 GCATATCTATCAAAAAAACTGGATCCGCTGGCCAGCGGGTGGCCAACTGCTGTAAAGCG 4150
QY 802 atcgacgtgtgacatctgttcaagaagcctgacaatgtactttggagacagaatata 861
Db 4151 GTTCGAGCAGTAGCACTCTCTCTCAAGAGCCTGATATAGTTAACTTTGGGCAAAATGTG 4210
QY 862 actgtaatagcccccatgtcatltagagaacatgctgcgaagcccccagacgaatgag 921
Db 4211 ACTGTGATGTGCTTCCATAGCCTCGAAAGCATCTGCGGCAACCCCGCAGCGGTGATG 4270
QY 922 accaagcccgcatgacccactatcaagcctgttctcacaagagaggtgcactttgcct 981
Db 4271 ACCAATGCGCAATAGACTCATTTACCAAGACCTGCTGTTAAATGAAGGGATTCCTTTGGC 4330
QY 982 ccaccagccgctctcaaccctgcaactctctgctgaagagactgtatgaacagtgact 1041
Db 4331 CCCCCGTGTCCTAAACCCAGCTACCTTCTCACTGCAAGTGGAGAACCCACCCAGTG 4390
QY 1042 catgatgtgcaatcaactatgtattgagagacgtgggtccggaagacacttaagacata 1101
Db 4391 CACAGGTGCTCAGAAATCTCTCGCGAAGAACTGGAACCTGACGAGACCTAGAGAACCA 4450
QY 1102 ccgctgtgctgagagatgtcactgttcaactgagcgaagcagctatgtgtgagagct 1161
Db 4451 CCAATTGCCCGG---GGTGCACACCTGGTATACAGACGCTAGCAGTTTCATCACGGAAGT 4507

QY 1162 aagagatgtgtggggcgagtgtagcgggagcccgacagatctlgcccaagcgtg 1221
Db 4508 AAGCGAGAGAGAGGAGGCCCGGATGCTAGATGGCAAGCGGAGGATGGGTACACGCTG 4567
QY 1222 ccggaaggaacttcagcgcaaaagcgtgagctcctatgcccctcagcaagcttgcggtc 1281
Db 4568 CCAGAGGTACGTACGCGCCAGAAAGCTGACATGATACCTTTGACGAGCGCATTAAGCTG 4627
QY 1282 gcggaagggaatccataaacttatacgaagcagcagtgatgctcttgacgtgacac 1341
Db 4628 GCCGAGAGAAAAAATCAATCAATCTACACGAGCAGAGATGCTTTGCCACTGCTCAT 4687
QY 1342 gtacagggggcactctataaacaaggggtgttctaccccaagggaggaagataag 1401
Db 4688 ATTATGGGCAATATATTAAGCAGAGGGGCTGCTCACTTCTGTGAAAAAGTATCAAA 4747
QY 1402 aacaaagaggaatctcctaagccttataagagccttaccatttgccaaaagcctagctat 1461
Db 4748 AACAAAGAGGAATTTTGGCCCTGCTAGAGGCCATCATCTCCCTAGGGCGGTGCCATT 4807
QY 1462 atacactgtccttgacatcaagaagcccaagatcatalctagagggaacagatgct 1521
Db 4808 ATCCACTGCTCTGGCCACAGAGGGGAATGACCTGTGGCCACTGGGAACCGAGAGGCC 4867
QY 1522 gaccgggtgccaagcagcagccagcgtgttaaccttgctctataatagaaagccc 1581
Db 4868 GACGAGGCTCAAAAGCAGCGCCCTCTTCCACCAAGAGTGTGGCAGGAA-----CTAC 4920
QY 1582 aaagcccccagaaaccagacaglacaccccttagaagacttgcgaagagataaaagata 1641
Db 4921 AAAACCTCAAGAGCAATTCAGGCCCGCTCAAGAAAGACAGGCCGAGGGAG----- 4972
QY 1642 gaccagttccttgagactccggaaggagacctgtacatcatalatggaagaaatcctg 1701
Db 4973 -----CTC 4975
QY 1702 ccccaagaagaggttagaatalgtccaacagatacatcgtctaaaccacttaggaact 1761
Db 4976 ACCCTGACCGGGGAAAGATTTCAATTAAGCGTTACATCAGTTAACTCACTTAGAGCA 5035
QY 1762 aaacacctgacagatgtgtcagaacatccctctacatgttcttgagcttaccagaagtg 1821
Db 5036 GAAAGGCTTCTCCACTAGTAGAACCGACCAAGCTCTCTCAACCACTCCAAATGTGA 5095
QY 1822 gctgactggtgtgtaacaactgtgtgcccctgcagcgtgttaatgtactaatccctcaga 1881
Db 5096 GTTCGGAAAGTCAACAGTCAAGTGTGAGGCTTGTGCCATGACTAATGCGGTCAACCACTAC 5155
QY 1882 ataccctcaggaagagactaaagggaagccacccagcgtctcactgggaagtgtgacttc 1941
Db 5156 AGAGAGACCGGAAAGAGCAAGCAGAGATGCAACCGGGGTGTACTGGAGAGTACCTTC 5215
QY 1942 actgagtaagcggcgtctaataatcaggaacaatactatctgtgttcttgagacactt 2001
Db 5216 ACAGAAATTAAGCGCTGTGTGTATGGAACCAAGTATCTTTAGTATCAATGATCTTC 5275
QY 2002 tcaagatgtgttagagcttatacctactaaagaagagacttcaacggtgtgctagaaga 2061
Db 5276 TCCGATGTGGTACAACTTTCTCTCAAAACAGTAACCGCCCTAATCTGCTGTAAAAA 5335
QY 2062 atactggaagaaatcttccaagatttgaaacttaagtgaaagtgaacatgtg 2121
Db 5336 ATATTGAGGAAGAAATTTTACCCCGCTTCGGGATTCCTTAAGGTACTCGGTACAGCAATGGC 5395
QY 2122 ccagcttctgttgcacagtagtcaaggaacttgcgaagataatgtggatltatggaaa 2181
Db 5396 CCGGCTTGTGTGCTCAGTAAGTACGAGGACGTGGCACTCACTCGGGGATTAATTGGAAG 5455
QY 2182 ctgcatltgtcatatagaccccaagctcaggaacaggttagagaagatgaatgaacatc 2241
Db 5456 TTACATTGTGGCTATAGACCCCAAGCTCAGGTGAGTAAGAAAGATGACAGAAATTT 5515

Query Match	28.18;	Score 933.6;	DB 2;	Length 8323;
Best Local Similarity	60.08;	Pred. No. 1.2e-273;		
Matches 1702; Conservative	0;	Mismatches 1094;	Indels 42;	Gaps 7

Db	3529	CTCTGGATCCCTGGGGTTTGGAGAAATGGCACACCCCTTTGATCCCTCTCACCAAAAGGGG	3581
QY	559	ggggagatctctcgtggctctctgaagcaccagaaggcattgatctatcaaaaagccctg	618
Db	3589	ACTCTGTTTGAATGGTGGGGCCCGAGACAGCAAAAGGCGCTACCAAGAGATCAAGCAGCGTC	3641
QY	619	ctgagagcaacctgctctggtccctccctgaagtaactataaccttaacctttatgtgat	678
Db	3649	TTTACTGCCCCCTGGCCCTTGGATTGGCCACACTTGTACTAAGCCCTTTCGAACCTTTTGTGAC	3701
QY	679	gagcgtlaaaggagtgatcccgaggaggttttaacccaaccccttaagacatagagagact	738
Db	3709	GAGAAAGCAGAGGCTAGCCCAAAAGGTCTCTTAACGCAAAAATCGGGCTTGTGGCTCGCGCG	3766
QY	739	gttgctactcgttcaaaagacttgatctctgtaagccagtggtgtggccgtatgtctgaag	798
Db	3769	GTGGCTACTCTTCACAAAAGCTTAGACCAATGGCACCTGGGTGGGCCCTTGCTCCTAGGG	3822
QY	799	gctatcgaagctgtgtgccatactgttcaagaagcgttcaaaattgactcttggagcaaat	858
Db	3829	ATGGTAGCAGGCGCATTCGCGGTCTTACCAAAAGACCTCTGGACAGCTCACCATATGGAGGCCA	3888
QY	859	ataactgttaataagcccccatgcatcttgagagaactgltcggagcccccagaccgatg	918
Db	3889	CTAAGTCATTCTGGCCCCCCTCAATGCACTGTGAAGGCACTATTAGCAACCCCTGTATGCTGTG	3946
QY	919	atgaccaaacgcccgatgatacccaatacxaagccgtctctc---caagaagagtgtaact	975
Db	3949	CTCTCCAAAGCCCGGATGACCCTACCAAGGCTGTGTTGTGGACAGGAGCGAGTCCAG	4006
QY	976	tctgcttcaacgaagccgctcttaacctgtcaacctctctcgtctgaagagactgatgacca	1035
Db	4009	TTTCGGACCAATATGTGGCCCTTAACCCGCTACGCTGTGCTCCCTCTA---CTGAGAGAGGGG	4065
QY	1036	gtgactactgtattgccatacctaattgatatggagagacttgggttcgcgaagacttaca	1095
Db	4066	CTGCACACTGTACTGCTTTGACATCTTGTGGTGAAGCCCAAGCAACTAGACCAATCTTACG	4122
QY	1096	gacataccgctgacttggagaagtgcttaacctgttcaactgaacgaagcagctatgtgtg	1155
Db	4126	GACCAAGCCTCTGCCAGAGCGTGACCAACCTGTGTACAGAGATGGAGAGAGCTTCTCTGCA	4185
QY	1156	gaaggtlaagagagatgctctggggcgagctgtgtgaacggaaacccgcagtcgtggccagc	1215
Db	4186	GAGGGGCGACGCCAAGGCGCGGAGCGAGCTAGTAACCAACGAGCGAGGTACTGTGGGCCAA	4245
QY	1216	agcctgcccgaagaactcttaagcgcacaagagctgagatcatgtgcccctcaagcagcttg	1275
Db	4246	GCACTGCCAGCGGGAGCATCGGCCCAAAAGCTGATGTATGTATGCGCTCAACCCAAACCTTA	4305
QY	1276	cggctggcggaaggaaatcatataaacttatagaagagacagatgatgtcccttgcgct	1335
Db	4306	AAATGGCAGAGAGGTAAAGACGTAATGTTTAAACCGATAGCGGTATGTCTTTCCTACT	4365
QY	1336	gcacacgtcacacggggccatclataaacaaggggtgtcttaacctcagcaagggagaa	1395
Db	4366	GCCCATATTTCACGGAGAAATATATAGAGGCGGGGTTTGCTCATCATCAGAAAGAAAGAA	4425
QY	1396	ataaagaacaaagagaattctlaagcctatattagaagccttaccattgtgccaaaagacta	1455
Db	4426	ATCAAAAATATAGACGAGATCTTGGCCCTTAAGTAAGGTCTCTTCTGCCCAAAAGACTT	4485
QY	1456	gctatataaactgtctctgtgaactcagaagccaaagatctatattctagaaggaaacag	1515
Db	4486	AGCAATAATTATTTCCTCCGGGACATATCAAGAGGAAACCGGCGGAGAGCAAGGGGCAACAG	4545
QY	1516	atgactacccgggttcttccaagcagagccagccagcgttgttaaccttctcctatatagaa	1575
Db	4546	ATGGCCGACCAAGCGGCCCGAGAGTAGGCCACTAGAGAACTCCAGAGACTTCCACACTT	4605
QY	1576	agcgcccaaaagcccgagaacc--agaagacagtagtacccttagaagactgtgcgaagataaa	1634

Db 4606 CTGATAGAAAATTCAGCCCCCTTATACATGACATTTTCTATACGGTGACTGACATA 4665
 Qy 1635 aaagatagaccagttctcttgagactc-----cgagaggagactcttaactcatat 1686
 Db 4666 AAAGTCTGACTTAAGCTAGGGGCCACTTATACATGCAAGAAAGTGTGGCTTATACAG 4725
 Qy 1687 gggagagaatctctgccccacagaagaggttagaataatgtccacagatatctgtcta 1746
 Db 4726 GGAAGAGCCTGTAAATGCTGATCAATTCACCTTTGACATATAGATTTTCTTCATCANTG 4785
 Qy 1747 acccaactag-----aactaaacactcgacagttgttcagaacatccctatcat 1800
 Db 4786 ACCCACTGACTTTCTCTCAAAAACAAAGGCTTCTAGAAAGAACTACTGCTTATTTAC 4845
 Qy 1801 gtctgaggtctaccaggagtgctgactcgggtgttcaaacattgtgacctcgacatg 1860
 Db 4846 ATGCTGAACCGGGATCGAAGCTTAAGACATCTACTGAGACTTCCCAAGCTGTGACAG 4905
 Qy 1861 gtaatgtctctcccaagaatacctccagaagaagagactlaagggagagccacccagc 1920
 Db 4906 GTCAATGCGACAGACTGCGCGTCAAAACAGGACTAGAGTTCAGGGGCAACCGCCGGC 4965
 Qy 1921 gctacgtggagagtgagacttactcagtgaggttaagccggtctaatacgaacaatatcta 1980
 Db 4966 ACCCACTGGGAAATTTGATTTCTACAGAGTAAACCTGCGCTGTATGGTATTAATATCTT 5025
 Qy 1981 ttggtttttgtagacaccttttcagagtgtagaggtctatctctactaagaagaagact 2040
 Db 5026 TTAGTTTTCATAGACACTTTCTCTGATGGGTAGAAAGCTTTCCCAACCAAGAAAGAACT 5085
 Qy 2041 tcaacgtgtgtgtctaaagaataactcgtgaggaanaatttccaaagatttgaataactaag 2100
 Db 5086 GCCAAAGTTGTAACCAAGAAAGCTACTAGAAAGAAATTTCCCGAGATTGCGCATGCCACAG 5145
 Qy 2101 gtaataggttcagacaaatgttccagcttctgttccaggttaagtcagagctgagcaag 2160
 Db 5146 GTATTGGGAACCGACATGGGCTGCTTCCCAAGGTAAAGTCAAGAGTCAAGAGTCAAGCGAT 5205
 Qy 2161 atattggagatttgaataactcgtgtgcatcagacagcccaagctcagaagagta 2220
 Db 5206 TTACTGGGGGTGATTGGAATCTGATTTGCTTACAGACCCCGAGCTTAAAGTCAAGTA 5265
 Qy 2221 gagaagatgaataacatthaagaagagacccctaacaaattgacacagagactgcat 2280
 Db 5266 GAAAGAAATGAATAGCAATCAAGAGAGACTTTAACTAAATTGACGCTTGCAACTGGCTCT 5325
 Qy 2281 aatattgagtgctctccgaccttgttctttagaggttgaaggaacacccctgtgacag 2340
 Db 5326 AGGAGCTGGGTCTCTGCTTCCCTGATGCGCTGATGAGCCCGCAACACGCGGGGCC 5385
 Qy 2341 tttagagtgacccctctatgattgctctacggggaaccccccgtgtgagagaattgac 2400
 Db 5386 CATGGTCTCAACCCCATGAAATCTTATATGAGGACACCCCGCCCTTGTAAACTTCCT 5445
 Qy 2401 ttgacacatagtgctgctgtgttccagaccttgttctctagtcagtcagagcgctc 2460
 Db 5446 GATCTGACATGGAAGAGTTTACATCAACCCCTCTCTCAAGCCCATTTACAGGACACTC 5505
 Qy 2461 gagggtgagagcgagcgaggttgaagcaactccggagagccctactcagaaggaagactg 2520
 Db 5506 TACCTGCTCAGACGAGGCTTGAGACCGTTGGCGGCACTTACCAAGAAACAATGAGAC 5565
 Qy 2521 c-----aagttcacactcgtctcaagttgagagatcagctatgtttaaagccacgct 2574
 Db 5566 CGGCGGCTAGTGCCTCAACCTTTCCGATGGTGACACATGTGGGTCCGCAACACACAA 5625
 Qy 2575 gcaggaacactcgagactcgggtgaaggaacctatctctactttgacacacacaag 2634
 Db 5626 ACTAAATATCTAGAAACCCCGCTGGAAGGACCTTATACGCTCTACTAGACTACCCCAAC 5685
 Qy 2635 gcttgaagtcgaaggaatcccaactgataatgcatcccaagcttaagcggc----- 2690
 Db 5686 GCTCTCAAGTGGAGGCAATTCAGCGTGGATCCACGCTGCCACGTAAGGCTGCCGAC 5745

Qy 2691 -----gccacctccagatcggggtggaagccgaaaagactgaatccctt 2739
 Db 5746 ACCAGATTGAGCCACCATGGGAATGACATGGCGGTTCACAGCTCTCAAAATCCCCCTA 5805
 Qy 2740 aagcttcgctccatcgcgtgtgttcttactctgtlcaataactcctcaagttatgtaa 2799
 Db 5806 AAGATTAAGATTGACCCCGGGGACCTCTATCCCTTAATTTCTCTCTGCTCTCAAG 5865
 Qy 2800 acgcttgtggaagacc 2817
 Db 5866 GGGCCAGATCCGACGAC 5883
 RESULT 9
 PCT-US93-08041-8
 ? Sequence 8, Application PC/TUS9308041
 ? GENERAL INFORMATION:
 ? APPLICANT: The Public Health Research Institute of the City of
 ? APPLICANT: New York, Inc.
 ? TITLE OF INVENTION: FUSION GLYCOPROTEINS
 ? NUMBER OF SEQUENCES: 16
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Davis Hoxie Faithfull and Hapgood
 ? STREET: 45 Rockefeller Pl.
 ? CITY: New York
 ? STATE: New York
 ? COUNTRY: USA
 ? ZIP: 10111
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patentin Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: PCT/US93/08041
 ? FILING DATE:
 ? CLASSIFICATION:
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 07/938,100
 ? FILING DATE: 28-AUG-1992
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US 08/_____
 ? FILING DATE: 20-AUG-1993
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Jacobs, Seth H.
 ? REGISTRATION NUMBER: 32,140
 ? REFERENCE/DOCKET NUMBER: 11698A50
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 212-757-2200
 ? TELEFAX: 212-586-1461
 ? INFORMATION FOR SEQ ID NO: 8:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 8323 base pairs
 ? TYPE: nucleic acid
 ? STRANDEDNESS: double
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: DNA (genomic)
 ? ANTI-SENSE: NO
 ? PCT-US93-08041-8
 Query Match 28.1%; Score 933.6; DB 5; Length 8323;
 Best Local Similarity 60.0%; Pred. No. 1.2e-273;
 Matches 1702; Conservative 0; Mismatches 1094; Indels 42; Gaps 7;
 Qy 19 acagatgcctctcttcgctgagattacacccacatgacacacacttctgcttga 78
 Db 3049 AAAGATGCTTTTCTGCTGAGACTCCACCCACAGTCACTCTCTTGGCCCTTTAG 3108
 Qy 79 tggagagatcaggtlaaggaagacgggagctacacttgagaccgactgcccagaag 138

Db 3707 CTAGTCACTCTGGCCCCCATGACAGTAGAGCACTAGTTAAGCAACCCCTGATCTGCG 3648
Qy 919 atbaaccaagcccgatgacccactatcaagccctctc---caagaagaggtcact 975
Db 3647 CTCTCCCAACCCCGAAGACCCACTACAGGCTCTGCTTGTGACGCGCCAGCCAGCCAG 3588
Qy 976 ttgcctcaccaagccgtctccaaacctcctctctgctgaagagactgtatgaca 1035
Db 3587 TTGCGACCAATATGAGCCCTTAACACGCTACGCTGCTCTCTA---CTGAGGAGGG 3531
Qy 1036 gtactcatgatgtgacatcaactatgtatgagagactgtgggtccgcaagagccttaca 1095
Db 3530 CTGCAACATGACTGCTTGTACATCTTGGCTGAAGCCGACGAGACTGACAGATCTTACG 3471
Qy 1096 gacataccgtctgactgtgagaagtgtcactgtgttactgtgaggaagcagctatgtgtgt 1155
Db 3470 GACCAAGCTCTCCAGAGCTGACACACCTGCTACACAGATGGAGCAGCTTCTGCA 3411
Qy 1156 gaaggtaaagagatgt 1215
Db 3410 GAGGCGACGCGCAAGCGCGGAGCAGCAGTACACCGAGAGCAGAGTACTGTGGCCAA 3351
Qy 1216 agcctgacggaaggaacttaagcgcaaaagctgaagctatgtgacctcaagcaagcttg 1275
Db 3350 GCACTGCGACCGGAGCATGCGCCCAAGAGCTGAGTGTAGCGCTACACCAAGCCTTA 3291
Qy 1276 cgagctgcggaaggaatcacaataattatagaagcaagagatgactgtgtgtgtgtgt 1335
Db 3290 AAATGCGAGAGGTAAAGAGCTGAATGTTACACCGATAGCGGTATGCTTTTGGCACT 3231
Qy 1336 gcaacagtaacaggggacatctataaacaagaggggtgtgtgtgtgtgtgtgtgtgtgtgt 1395
Db 3230 GCCATATTCACGAGAAATATATAGAGGCGGGTGTGTACATCAGAGAGAGAAAGAA 3171
Qy 1396 ataaagaacaagagaatcttaagccttataagaagccttactatgtgtgtgtgtgtgtgt 1455
Db 3170 ATCAAAATTAAGGCGAGATCTTGGCCCTTACTGAGAGCTCTCTTCTGCCCAAAACAT 3111
Qy 1456 gctatitacactgtctctgacatcagaagaagccaagatctatctatctagaagacag 1515
Db 3110 AGCATATTAATGCTCCCGGAGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3051
Qy 1516 atggtctacacgggt 1575
Db 3050 ATGGCCGACCAAGCGCGCCAGAGATGACACTAGAGAAACCTCCAGACTTCCACACT 2991
Qy 1576 acgcccacaagcccaagaagcc-agaagacagtacacccctagaagactgtgcaagaagataa 1634
Db 2990 CTGATAGAAAATTCAGCCCTTACTACATGAACATTTTCACTACGCGTGAACATA 2931
Qy 1635 aaagatagaccaggtctctgt 1686
Db 2930 AAAGATGCTGACTAATAGAGGCGCCACTTATGAGATGCAAGAGAGAGAGAGAGAGAGAG 2871
Qy 1687 gggagaagaatctctgt 1746
Db 2870 GGAAGGCTGATTAATGCTGATTAATGCTGATTAATGCTGATTAATGCTGATTAATG 2811
Qy 1747 acccaactag-----aactaaacctgtgcaagagtggtgtgtgtgtgtgtgtgtgtgtgtgt 1800
Db 2810 ACCCACTGAGTTTCTCAAAAACAAAGGCTTCTTAGAAAGAACTACTGCTTATATAC 2751
Qy 1801 gtctgtgagctacaggaagtggt 1860
Db 2750 ATGCTGAACCGGATGCAAGGCTCAAGAGATCACTGAGACTTCCAGAGCTGTGACAG 2691
Qy 1861 gtaatgtcaatctctcagaatactcctcagaagaagactaaagagagcaagcccaagc 1920
Db 2690 GTCAATGCTCAAGAGCTGCGCTCAAAACAGGAGCTAGAGTGTGAGGCGACCGACCGCG 2631
Qy 1921 gctcactgtggaagt 1980

Db 2630 ACCCACTGGAAATGATTTCACTGAGGTAAACCTGGCTGTATGGGTATATATCTT 2571
Qy 1981 ttggtttttgtagacaccttctcaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2040
Db 2570 TTAGTTTTCATAGACACTTCTCTGTATGGGTGTAGAGCTTTTCCACCAAGAGAGAACT 2511
Qy 2041 tcaacctgt 2100
Db 2510 GCCAAAGTTGTAAACCAAGAGCTACTAGAGAAATCTTCCCAATTCGCGATTCGCGATCCAG 2451
Qy 2101 gtaataaggtcaagaacatgt 2160
Db 2450 GTATTGGGAGACCGCAATGGGCTGTGCTTCTTCCAAAGGTAGTCAACAGTGTAGCGAT 2391
Qy 2161 atattgt 2220
Db 2390 TTATTGGGGGTGTATTGGAAACTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2331
Qy 2221 gaaagagatgaatagaacattaaagaaccttaccacaaattgacacagaagactgtgcaat 2280
Db 2330 GAAAGATGAATATAGACATCAAGAGACTTAACTAATTAAGCGCTTGAACCTGGCTCT 2271
Qy 2281 aatgtatgt 2340
Db 2270 AAGGAGCTGGGTCT 2211
Qy 2341 ttgt 2400
Db 2210 CACGCTTCACCCCATATGAAATCTTATATGGGAGACCCCGCCCTGTAACTGCTCT 2151
Qy 2401 ttgtacacatgt 2460
Db 2150 GATCTGACATGCAAGAGTACTCTATACCCCTCTCTCAAGCTCATTTACAGGACATC 2091
Qy 2461 gagt 2514
Db 2090 TACCTGTACAG 2031
Qy 2515 gactgtcaagatctcaccatogcttccaaagtgtgagatltcagtlcgtltgtgtgtgtgtgtgt 2574
Db 2030 CGGCGGATAGTACCTCACCCCTTCCGGGTGCGTGACAGTGTGGTCCGAGACACCA 1971
Qy 2575 gcaagaacactgtgaactcgt 2634
Db 1970 ACTAAAAATCTAAGAACCCCGCTGTGAAGAGACCTTATCCGTCCTACTACTACCCCCAC 1911
Qy 2635 gctgt 2690
Db 1910 GCTCTAAGATGAG 1851
Qy 2691 -----gcacactccgattcgt 2739
Db 1850 ACCAGAGATGAGCCACAG 1791
Qy 2740 aagctgt 2799
Db 1790 AAGATTAAGATTAACCGCGGAGCTCTTATCCCTTATTTCTCTCTGTCTCAAG 1731
Qy 2800 acgctgt 2817
Db 1730 GGGCCAGATCCGACGAC 1713

RESULT 11
US-08-886-642-9/c
; Sequence 9, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:

Db 2990 CTGATAGAAAATTCAGCCCCCTTACTGATGACATTTTCTACTATAGCGTACTGACATA 2931
 QY 1635 aaagatagaccagctctctgagctc-----cggaggggacctgctatactcatat 1666
 Db 2930 AAAGATCTGACTAACTAGGGGCCACTTATACACATGCAAGAGTGTGGGTTTATCAG 2871
 QY 1667 gggaaagaaatccctgcccacaaagaggttagaataatgccaagataacatcgtcta 1746
 Db 2870 GGAAGAGCTGTAATGCGCTGATCATATACCTTTGACATTTAGATTTTCTTCATCATATG 2811
 QY 1747 accacactaag-----aactaaacactcagcagctgtgtcagaacatccctatcat 1800
 Db 2810 ACCCACTGCTCACTTTCTCTCAAAAACAAGGCTCTTCTAGAAAAGAACTAGCTGCTTATATAC 2751
 QY 1801 gttctgaagctaccagagagtgctgactcggtggttcaaacatgtgtgacctgcccagctg 1660
 Db 2750 ATGCTGAACCGGGATTCAGACCTCTAAAGACATCTAGACTTGTCCAAAGCTGTGCACAG 2691
 QY 1861 gtaattgtaactcctccagaatacctccagaagaagactaaaggagagccacccagcgc 1920
 Db 2690 GTCAATGCGCAGCAAGTCTGCGCTCAACAAGGGACTAGAGTTGAGGGGACCCAGCCGGC 2631
 QY 1921 gctcactgggaagtgagagcttcacgtcagtgtaagccggtcctaatacaggaacaatatcta 1980
 Db 2630 ACCCACTGGGAATTTGATTTCTACTGAGGTAAACCTGGCCTGATGGGTAAATATCTT 2571
 QY 1981 ttggttttttagacacactttccagatggttagagagctatctactactaaagaagagct 2040
 Db 2570 TTAGTTTTCATAGACATTTCTCTGATGGGTAGAACTTTCCCAACCAAGAAAGAACT 2511
 QY 2041 tcaacctgtgtgctaaagaataactcggaggaatattttccaaagatttgaatacctaag 2100
 Db 2510 GCCAAATTTTAAACCAAGAGCTACTAGAAATATCTTCCCGCAGATTCCGCGATGCCACAG 2451
 QY 2101 gtaataagctcagaacaatgttccagcttctggtcccaagtgtaagtgagagcgcgcag 2160
 Db 2450 GTATTGGGAGCCAGACATGGGCTGCTGCTTCCCAAGTAAAGTACAGACATGACCGCAT 2391
 QY 2161 atattggagattgattggaactgcatgtgtacatacagaccacaaagctcagagacagata 2220
 Db 2390 TTATTGGGGCTTATGGAATACTACATTGTGCTTACAGACCCAGAGTTCAAGTCAAGTA 2331
 QY 2221 gagaagatagataagacatlaaagagaccttaacaaattgacacagagagctgcat 2280
 Db 2330 GAAGAAATGATATGGAATCAACAGAGACTTAACTAAATGACGCTTCCACTGCTCT 2271
 QY 2281 aatgatagatgtctcctcctgcttctgcttcttaaggtgagagacaccccttgacag 2340
 Db 2270 AGGAGACTGGGTGCTCTGCTTCCCTTGGCTTATCAGACCCGCAACAGCCGGGCCCC 2211
 QY 2341 ttggagctgaacccctcatgaattgtctacaggggagaccccccctgttgacagaatgtcc 2400
 Db 2210 CAGGGCTCTACCCCATATGAATCTTATATGGGGGACCCCGCCCTTGTAAACTTCCCT 2151
 QY 2401 ttggacataatgtgtatgtgctgcttccacgcttcttctcaggtcagagcgcgc 2460
 Db 2150 GATCCTGACATGGCAAAAGTTACTCTAAACCCCTCTCTCAAGCTCATTTACAGAGCATC 2091
 QY 2461 gaagtggtagagcagcagcgtlgaagcagctccggagagcctactcag-----agga 2514
 Db 2090 TACCTGGTCCAGCAGCAAGTGTGAGACCGTGGCGGAGCTTACCAAGACGACTAGAC 2031
 QY 2515 gacttgcaagttccaatcgcctccaaglttgagattcaagltatgttagagcagcagct 2574
 Db 2030 CGGCGGTGTACTCTACCTTTTCCGGGTGGTGACACAGTGTGGGTCCGCGAGACAA 1971
 QY 2575 gcaagaaacctcgaagactcgtgtggaagagacttctcgtacttttgaccacacacag 2634
 Db 1970 ACTAAAAATCTTAGAACCCCGCTGGAAAGAACCTTATACCGTCTACTAGTACCCCAAC 1911
 QY 2635 gctgtgaaagctcgaagaaatcccaactgatacgtcagtaacagtaacagcgcgcgcgc 2690
 Db 1910 GCTCTCAAGTATGACGGCATTCACAGGTGATCCACGCTGCCACGTAAGGCTGCGCAGC 1851

QY 2691 -----gccacctccagattcgggttggaaagccgaaagacttgaataatccctt 2739
 Db 1850 ACCAGGATTTGAGCCACAGCAAGATGACATGGCGGTTCACAGCGTCCCAAAATCCCTTA 1791
 QY 2740 aagctgcctccatcagcgtgtgttcttctcgtactcgttaataactcctcaagtaagta 2799
 Db 1790 AATATAGATTGACCCCGGGAGACTTCTATCCCTTAATTTCTCTGCTCTCAAG 1731
 QY 2800 acgccttltggaagccc 2817
 Db 1730 GGGCGAGATCCGACGAC 1713

RESULT 12
 PCT-US93-08041-9/c
 ; Sequence 9, Application PC/TUS9308041
 ; GENERAL INFORMATION:

; APPLICANT: The Public Health Research Institute of the City of
 ; APPLICANT: New York, Inc.
 ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
 ; STREET: 45 Rockefeller Pl.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10111

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/08041
 ; FILING DATE:

; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/938,100
 ; FILING DATE: 28-AUG-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/_____
 ; FILING DATE: 20-AUG-1993

; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jacobs, Seth H.
 ; REGISTRATION NUMBER: 32,140
 ; REFERENCE/DOCKET NUMBER: 11698A50

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-757-2200
 ; TELEFAX: 212-586-1461

; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 10367 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO

PCT-US93-08041-9

Query Match 28.1%; Score 932; DB 5; Length 10367;
 Best Local Similarity 59.9%; Pred. No. 4.3e-273;
 Matches 1701; Conservative 0; Mismatches 1095; Indels 42; Gaps 7;

QY 19 acagatgcctctctcgtcgtgagattacacccacagcaccactttgcttcgaa 78
 Db 4547 AAAAGATGCTTTTCTGCTGAGACTCCACCCAGTCAAGTCTCTTCGCTTTGAG 4488
 QY 79 tggagagatccaggttacgggaagaccgggacgtcactcagaccgactgcccagaagg 138

Dh 4487 TGGAGAGATCCAGAGATGGGAATCTCAGACAAATTAACTGGACCAAGACTCCCGAGGAT 4428
Qy 139 ttaaaagactcccccatctttgacgaagccctacacaggaacttggccaacttcaag 198
Dh 4427 TTCAAAACAGTCCACCCCTGTTGATGAAGCCCTGCACAGGAGACTCGACAACTTCGG 4368
Qy 199 atcaaacacctgaagttagacctctccagtagacgtgtagtacccttcttggcggaagc 258
Dh 4367 ATCCAGACCCAGACCTGATTTGCTCCAGTATGTAGTACTACTGCTGGCCGCACT 4308
Qy 259 accaaacagactgttagaaggtacgaaggaactacttgcctgaattgtcctacgaagc 318
Dh 4307 TCTAGCTTGAATGTCACAAAGATGCGGGCCCTGTTACAAACCTTAGGGAACTCGGA 4248
Qy 319 taacagacctctgttaagaagccagatttgcagagaagagtagtaacatacttgggtgac 378
Dh 4247 TATGGGCTTCGGCCAGAAAGGCCAAATTTGCCAGAAAGGTCAAGTATCTGGGGTAT 4188
Qy 379 agtttcggggcgagcgatggtcgtagcggaagcaaggaagaactttagtccagata 438
Dh 4187 CTTCAAAAGAGGGTCAAGATGCTGACTGACGCCAGAAAGAGACTGTGATGGGGCAG 4128
Qy 439 ccggcccccaacagcaagcaagtagagagtttttgggaacgtgatttgcaga 498
Dh 4127 CCTACTCGAAGACCCCTCGACAACTAAGGAGTTCCTAGGAGCGGAGGCTTCTGTGCG 4068
Qy 499 ctgtgatcccggttgcagaccttagcagccccactaccogctaaccaagaagaaaa 558
Dh 4067 CTCTGATCCCTGGGTTTGAGAAATGGCAGCCCTTGTACCTTCACCAAAAGGGG 4008
Qy 559 gggggaattctctgtgtctccagacacagaagcatttgaatgatacaaaagggccgt 618
Dh 4007 ACTCTGTTGAATGGGGCCAGACGAGAAAGGCTTACCAAGATGATCAAGCGCTCTC 3948
Qy 619 ctgagcgacctgtctgtgacctccctgaagtaactaaaccttaccctttagttagat 678
Dh 3947 TTAAGTGGCCCTGCTCCCTGGATGTGCAAGCTTGAAGCCCTTGAAGCTTTTGTGTAC 3888
Qy 679 gagcgtlaaggaagtagcccgaggaattttaaaccacacttaggaacatgagagagact 738
Dh 3887 GAGAAGCAGGGCTACGCCAAAGGTGTCCTAACGCAAAAACGTGGGGCTTGGGCGCG 3828
Qy 739 gttgctactctgtaagaagcttgaatccttgaagcttggccgtatgtctaaag 788
Dh 3827 GTGGCGCTACTCTGCAAAAAGTAGAACCAAGTGGCAGCTGGGTGGCCCTTGGCTACGG 3768
Qy 799 gctatcgagctgtgccaactactgttcaagagcgtgaacaaatgacttgggaagaaat 858
Dh 3767 ATGCTAGCAGCCATGCGCTTCTGACCAAAAGCGCTGGCAAGCTCACCATGGACGCCA 3708
Qy 859 ataactgtatagcccccaatgcatgtgagaacatcgttggcagcccccgagacagatg 918
Dh 3707 CTAGTATCTGCGCCCATGACAGTACAGGACATGTTAAAGCAACCCCTGATGCTGG 3648
Qy 919 ataacagacccccgaatgacacatacaaaactgtcttct---cacagaaggttcaact 975
Dh 3647 CTCTCCAGGCCGGAATGACCAACCAACGAGCTGCTGTGACACAGGACGAGTCCAG 3588
Qy 976 ttgcctcaccaagcagctcacaacctgtccactcttgcctggaagaagatgtagaaca 1035
Dh 3587 TTGCGACCAATATGCGCTTAAACCAAGCTACGCTGCTCCCTCTA---CTGAGGAGGGG 3531
Qy 1036 gtagactcatgattgcatcaactatgattgtaggaagactgggtgcgcgaagaccttaca 1095
Dh 3530 CTGCAACATGACTGCTTGAACATCTTGCTGTAAGGACCCAGGAACATGACATCTTACG 3471
Qy 1096 gacatacgcctgactggaagaatgtctaactgttcaactgacgaagcagctatgtgtg 1155
Dh 3470 GACACAGCTCTCCAGACGCTGACACACACTGCTGACAGATGGGAGCGCTTCTGTGAA 3411
Qy 1156 gaaggttaaggaatgagctgtggcggaagtgtgtaggaagccgcgaatgttggcagc 1215
Dh 3410 GAGGGGAGCGCAAGGCGGAGCAAGTACCAACCGAAGCCAGATGATCTTGGGCGCAA 3351

Qy 1216 agcctgcggaaggaacttcaagcgcaaaagcttgaatcatalgaccttcaacgaagcttg 1275
Dh 3350 GCACCTGCCAGCCGGACATCGGCCCAAGAGCTGAGTTATGAGCTCACCAACCTCTTA 3291
Qy 1276 cggctggcggaaggaatccataaacttatacgaagaaggaatgagcttgcagact 1335
Dh 3290 AAATATGCAACAGTAAAGCTGAATGTTACACGATAGCGGTAATGCTTTTCCACT 3231
Qy 1336 gcaacgttaacggggccatctatacaaaagggttgccttaccctcagcaggaagga 1395
Dh 3230 GCCCATTTCCGAGAAATATATAGAAAGCGCGGGTGTCTCATCATCAAGAAAGAA 3171
Qy 1396 ataaagacaagagaatcttaagcctatitagaagccttaacatttccaaaaggcta 1455
Dh 3170 ATCAAAATATAGAGACGAGATCTTGCCCTACTGAAAGCTCTCTCTGCCCCAAAGACTT 3111
Qy 1456 gctatatacactgtctgtgacatcgaagaagccaagatctatcttagaggaagcaag 1515
Dh 3110 AGCATTAATTTGCTCCGGGACATCAGAAAGGAAACCGCGGAGGCAAGGGGCAACAG 3051
Qy 1516 atgcttgaacgggttgcgaagcagcgccagcgttgaaccttgcctataatagaa 1575
Dh 3050 ATGGCCGACCAAGCGCGCCGAGAAAGTACCACTAGAAAGAACTCCAGAGACTTCCACACTT 2991
Qy 1576 acgccaaagccccagaagcc-agaagacgtacacccctgaagacttgcgaagagataa 1634
Dh 2990 CTGATAGAAATTTGACGCCCTTATCTATCATGAACTTTTCTACTAGGAGTGCATGATA 2931
Qy 1635 aaagatagacagcttctctgaagctc-----cggaaggagccgtgatacctatat 1686
Dh 2930 AAAGATCTGACTTAACCTAGGGGCCACTTATGACATGCAAAAGAAAGTCTGCTTATTCAG 2871
Qy 1687 ggaagaagaaatccgtgcccccaaaagaggtttagaatalgtlcaaacagatatactgtcta 1746
Dh 2870 GGAAGGCTGTATGCTGATCAATTCATCAATTCACCTTGAACATATAGATTTCTTCAATTA 2811
Qy 1747 acccaactaag-----aactaaaccttgagaagcttgcgaataatccctatcat 1800
Dh 2810 ACCCACTGAGTTTCTCAAAAACAAAGGCTTCTTCAAAAGAAAGTCTGCTTATTTAC 2751
Qy 1801 gttctgaggttcaacagaggttgcgtgactcgtgtgttcaaacatttgccttgcagctg 1860
Dh 2750 ATGCTGAACCGGGATGCAAGCTCAAAAGACTCACTAGATGATTCCAAGCCTGCGACAG 2691
Qy 1861 gttaatgttaactccttcaagaatactcccaagaaagaaactaaggggaagcccaagc 1920
Dh 2690 GTCAATGCCAGCAAGCTGCGCTCAAAACAGAGGACTTGAAGTTGCAAGGGCACCGCGCG 2631
Qy 1921 gctcaactgggaagtgagcttcaactgaagtaagccgggttaatacgaagaacaaatcata 1980
Dh 2630 ACCCACTGGGAAATGATTTCACTGAGGTAAACCTGGCTGTATGGGTATTAATATCTT 2571
Qy 1981 ttggttttgtgaacaccttccagatggttgaagcttactcctaagaagaagact 2040
Dh 2570 TTAGTTTTCATGACACTTTCTCTGATGAGTAAGCTTCCCAACCAAGAAAGAACT 2511
Qy 2041 tcaaccgtgtgtgcttaagaanaatacttgaaggaatlttccaaggttggtaacctaaag 2100
Dh 2510 GCCAAAGTTGTAAACCAAGAGCTACTAGAGAAATCTTCCAGATTTGGCATGGCACAG 2451
Qy 2101 gtaataggttcagacaatgtgtccagcttgcgttgcgtgaaggtlaagtcaggaagcttgc 2160
Dh 2450 GTATTGGAAACGACAAATGGGCTGCTGCTGCTCCAAAGTAAAGTAAAGTAAAGTAAAG 2391
Qy 2161 atattgggaattgattggaactgtatgtgcatatacagaagcccaagctcaggaaggtta 2220
Dh 2290 TTTATTGGGGTGTATTGGAATACATGTTGCTTACAGACCCCAAGATTTAGAGTAAAGTAA 2331
Qy 2221 gagaagatgataagccatitaaagagaccccttaacaaatitgaacacagagagacttga 2280
Dh 2230 GAAAGATGATGATGAGCAATCAAGGAGACTTTTAACATAATTTGACCGCTTGCAACGTGCT 2271

Dh 6217 agaataagatagaaccatcaagaagacttaactaataactgctgaactgctctaga 6276
Qy 2284 gatgagatgctctccgccccttctgcttcttaaggtaggaaccccttgacaagttt 2343
Dh 6277 gactggtgtctcctactccctcttagcccttgaccgaagcccgacagccggcccccac 6336
Qy 2344 gggctggccccccttaattgctctacgggggaccccccccgcttgggaatctgctt 2403
Dh 6337 ggcctcaaccacatagatgactatataatgggaccccccccgcttgaatctcctgac 6396
Qy 2404 gacatagtgctgtagtctgctctcccaagccttcttctctagctcgaagcgctcga 2463
Dh 6397 cctgacatgacaagaagtacttaacagccctctctcaagctcaacttaacagctctctac 6456
Qy 2464 tgggtgagcagcagcgctgtaggaacagctccgggaagccttaacaggaagagacttgcaa 2523
Dh 6457 ttagtcacagcagcagcagctctgagaccccttgcgagccttacaagaacacttgacga 6516
Qy 2524 -----gttccacatcgctcccaagttggagatcaatctatgttgaagcgcacgtgca 2577
Dh 6517 ccggtgtactcctcaacccttaaccgagctcgagacacagtggtgtccgcgacacagact 6576
Qy 2578 ggaacctcgagactcggtggaaggacactatctcgtactttgaccacacacagcgct 2637
Dh 6577 aagaacctagacccctcgctggaaggacacttaacacagctcctgctgacaccccccacgc 6636
Qy 2638 gtgaagtgcgaagaatcccaactgtagtcaatcagatccaggttaagccggc 2690
Dh 6637 ctcaaatgagcagcagcagctgtagtatacagcgcgcacagtgaaagctgc 6689

RESULT 15

US-09-011-745-2
: Sequence 2, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A
: APPLICANT: Takeuchi, Yasuhiro
: APPLICANT: Cosset, Francois-Loic
: TITLE OF INVENTION: Expression systems
: FILE REFERENCE: 09/011,745
: CURRENT APPLICATION NUMBER: US/09/011,745
: EARLIER FILING DATE: 1998-06-22
: EARLIER APPLICATION NUMBER: PCN/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 7616
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Portion of
: OTHER INFORMATION: construct
US-09-011-745-2

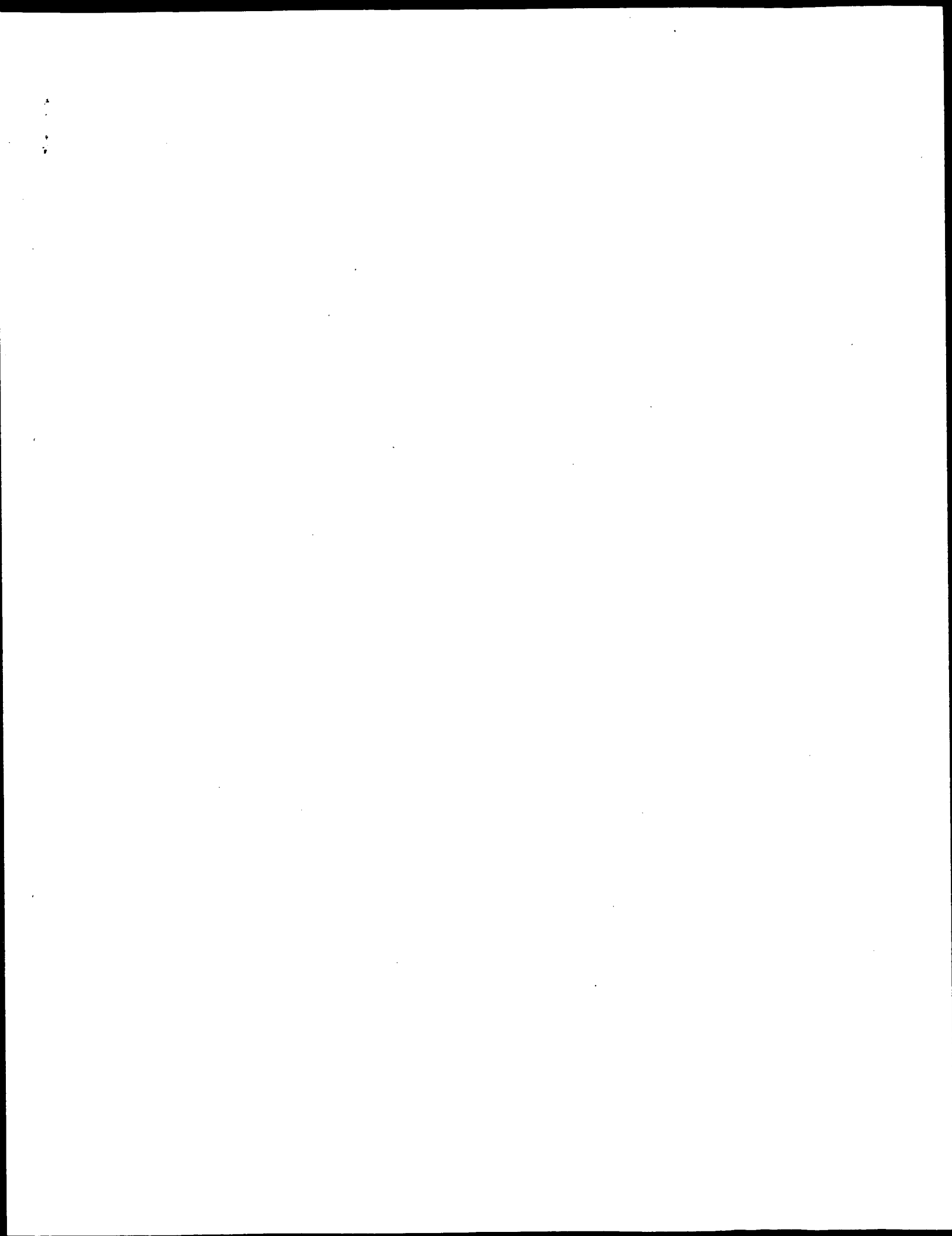
Query Match 27.6%; Score 916.2; DB 4; Length 7616;
Best Local Similarity 60.6%; Pred. No. 2.3e-268;
Matches 1633; Conservative 0; Mismatches 1033; Indels 27; Gaps 7;
Qy 22 gatgcctctctctgctgagatatacaccacactagccaacacacttttgccttcogaatg 81
Dh 3161 gatgccttctctgctgagatatacaccacacactagccaacacacttttgccttcogaatg 3220
Qy 82 agaatacaggtacgggaagaacacggcgagctacccctggccacgacgcccccaaggctt 141
Dh 3221 agaatccagatgaggaatctcctcaggaactgacccctggacacacgcccccaaggctt 3280
Qy 142 aagaactccacacatctttagcagaagccctacacagggacctgacacacttcagatc 201

Dh 3281 aaaaacagtcacacccctgtttagtaggcactgcacagacactagacacttccgagtc 3340
Qy 202 caaacacctcaggtgacccctctctcagtagcagtgagactctctctgcggaagcacc 261
Dh 3341 cagcacacagactgtatctctgctacagtagcgtgagtagtactctctgctgcgcactct 3400
Qy 262 aacacgactgcttgaagaagtagcagaagcactacgcctggaattgtctgacctagctac 321
Dh 3401 gagctagactgcacaaagatcctcggtccctgttacaacacccctgggaactcgtgact 3460
Qy 322 agagcctctgctcaagaagcccaagatttgcaggaagagatacacttgggttagact 381
Dh 3461 cgggctctgcaagaagcccaatttgcagaaacaggtcaagtatcttgggtatctt 3520
Qy 382 tgcggtgagcagcagcagctgctgacgagagcagcaggaagaacactgtgctcagtagcc 441
Dh 3521 ctcaaaagaggtcttagagatgagctgactgagcccaaaagagactgtgagtgagcact 3580
Qy 442 gcccaacacacagcacaacagtgagagagcttctgggacagcctgatttgcagactg 501
Dh 3581 actcogaagaacccctcgaacactaagggagctcctagaggagcagcgtctgtcgcctc 3640
Qy 502 tggatcccggtgttgcagacttagcagcccaactacccgttaacccaagaagaagg 561
Dh 3641 tggatccctgggttgcagaaatgagcccccctgtgacccctcctcaacaaaacgggact 3700
Qy 562 ggaattctctggtctctcgtgacacacaggaagatttgaactataaaaggccctgtg 621
Dh 3701 ctgttaatttggggcccaacacaaagagccttcaagaataacgaacactctctca 3760
Qy 622 agcgacactgctctgcccctccctcagcgaactaacccttaactatgtatgagtag 681
Dh 3761 actgcccacagccctgggtgttgcagaattgactaagcccttggactcttctgtagcag 3820
Qy 682 cgtaaaggagtagcccgaggaatttaaccacacacttaggaacacatggaaggaagactgt 741
Dh 3821 aagcagaggtacgcacaaagtgcttcaacgcaaaacttggacacttgcgtgcgcgtgtg 3880
Qy 742 gcttactcttcaaaagagctttagcctgtgagcaggtgttgcgcgtgacttgaagct 801
Dh 3881 gcttactcttcaaaagagctttagcctgtgagcaggtgttgcgcgtgacttgaagct 3940
Qy 802 atcgacgctgtgagccatctgttcaagagcgttgaacaaatltgacttggagacagata 861
Dh 3941 gttagcgcacattgctgacttgaacaaagatgcaaggaacttaacatgtaggaagcacta 4000
Qy 862 actgtaataagcccccaatgcatgtggaacacatcgttgcgacgcccccagacagatgagt 921
Dh 4001 gtcatctgtgcccccatgtagtagagcactagtagcaaacacccccagacgctgtgt 4060
Qy 922 accaagcccgatgagcccaactataaagcgtcttct--cagagagaggttactctc 978
Dh 4061 tccaagcccgatgagcccaactataaagcgtcttcttggagacaggaacgggttccagctc 4120
Qy 979 gctccacacgcccctcacaaccctgacactcttctgctggaagagactgatatcaacagt 1038
Dh 4121 ggaacggtgtgagccttgaaccccggttagcgtctcc--actgcttgaaggaaggtctg 4177
Qy 1039 actcatgattgcataactatgacttgaaggaagcactgggttccgcaagaacttaacag 1098
Dh 4178 caacacaactgcttgaatctcctgscggaagcccaagcagacccgacacttaacagac 4237
Qy 1099 ataccgctgacaggaagaatgtcctaacttggcttgcaggaagcagcactgtgtgtaa 1158
Dh 4238 cagcgtctcccaagcgcgaacacacactgttgaacaggaatggaagcagctcttcaacaag 4297
Qy 1159 ggtcaagagatgctgtggtgagcaggtgtagcaggaacccgacacatctggtccagcagc 1218
Dh 4298 ggaacagcgttaagcgtgagctgtgtagcaccaggaacccgaggtatctggtctaaagcc 4357
Qy 1219 ctgcggaaggaactcaagcgaagaagcgtgactatgagccctcaacgcaagcttgcgg 1278

Db 4358 ctgccaagcggaacatcgctcaagcggtgctgaatgataagcctcaacccagccctaaag 4417
QY 1279 ctgccaagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1338
Db 4418 atgccaagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4477
QY 1339 caagtcacgagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1398
Db 4478 catatcagtcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4537
QY 1399 aagaacaagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1458
Db 4538 aaaaaaagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4597
QY 1459 atatacagtcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1518
Db 4598 atatacagtcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4657
QY 1519 gctgacgagcggttgcgaagcagcggaagcggtatgctcttgcagctga 1577
Db 4658 gctgacgagcggttgcgaagcagcggaagcggtatgctcttgcagctga 4717
QY 1578 gcccaagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1637
Db 4718 atagaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4777
QY 1638 gataagc-----agtcctggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1689
Db 4778 gataagc-----agtcctggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4837
QY 1690 aaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1749
Db 4838 aaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4897
QY 1750 caagtcacggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1808
Db 4898 caagtcacggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 4957
QY 1809 gatacagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1868
Db 4958 gatacagcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5017
QY 1869 taatcctcaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1923
Db 5018 taatcctcaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5077
QY 1924 caagtcacggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 1983
Db 5078 caagtcacggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5137
QY 1984 gttttgtagaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2043
Db 5138 gttttgtagaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5197
QY 2044 accgtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2103
Db 5198 accgtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5257
QY 2104 ataggtggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2163
Db 5258 ataggtggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5317
QY 2164 ttggtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2223
Db 5318 ttggtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5377
QY 2224 aggaatggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2283
Db 5378 aggaatggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5437
QY 2284 gattggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2343
Db 5438 gattggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5497

QY 2344 ggcctgaccccatatgatactgctcaacggaagcaggtatgctcttgcagctga 2403
Db 5498 ggcctgaccccatatgatactgctcaacggaagcaggtatgctcttgcagctga 5557
QY 2404 ggcctgaccccatatgatactgctcaacggaagcaggtatgctcttgcagctga 2463
Db 5558 ggcctgaccccatatgatactgctcaacggaagcaggtatgctcttgcagctga 5617
QY 2464 tgggtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2523
Db 5618 tgggtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5677
QY 2524 -----gttcacacggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2577
Db 5678 ccggtggtggttaaggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5737
QY 2578 ggaacatcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2637
Db 5738 ggaacatcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5797
QY 2638 gtaagtcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 2690
Db 5798 gtaagtcggaacatcgctcaacatcttaacggaagcaggtatgctcttgcagctga 5850

Search completed: February 24, 2002, 02:39:30
Job time: 20121 sec



Db 3 GTGGTGTACGACTGTGGGCCCCAGCGCGCTTGGAAATAAATCCTTGTCTTTCATC 62
Qy 61 aagaccgtcttcgtgaagtgatt-tygggtgcgcgtctctccagaccgagagggga 119
Db 63 AAGACCGCTTCTCTGTGAGGATTAAGGGAGTCCCTTTTCGACCGCCAGAG----- 115
Qy 120 ttgtcttcttactggtccttcatcttggtgcgttgccggaaatccctccgaccacctt 179
Db 116 --TTCTTTTGTGCTTGTACATTGGGGGCTCGTCCGGATCTGTCCGGCCACCCCTA 173
Qy 180 aacaccggagaaacgaacttggagta-aaggatcccccttggaaatgltg--tggtc 236
Db 174 AACCCGGAGAAACGACTTGGAGTAAAGATCCTTTTAACTGATGATGATGATC 233
Qy 237 gcccggcgtctctgtctgaagtgtctgttcttcgtgtagcgcgtcttcggttcaactg 296
Db 234 GCGCCGCGCTCTGTCTGTGAGTGTCTTTCAGTGTGCGCGCTTTCGCTTTCAGCTG 293
Qy 297 tccctcagacgctlaagaaacttggagactgtgatacagaacgtcttagagagatcag 356
Db 294 TCCCTCAGCGCGTAAAGGCTGGGGGACTGTGATCAGACAGCTGCTAAGAGATCACAG 353
Qy 357 gctgcacaccttgggggagacgcccggaggatggggagagccagggagcgcgtgtgtctcc 416
Db 354 GCTGCGCGCTTGGGGGAGCGCCCGGAGGTAGAGAGAGCCAGGAGCGCTGTGTCTCC 413
Qy 417 tactgtcgtcagagagagacgagctctgttfttgaagcgaaagctccccctccggcgcg 476
Db 414 TACGTGCGTCAAGAGACCGAATCTGTGCTGAAGGAAAGCTTCCCTCCCGCGACG 473
Qy 477 tccgaactcttctgcgtctgttgaagaacgagcaggtgcgtgtgtctcgtatctgtg 536
Db 474 TCCACACTTTTGGCTGCTGTGTGAAGACGAGCGGCTCACGTTGTCTGATCTGTG 533
Qy 537 gttctcgtctcgtgtgtcttctgtctgtgtctgtctgtctcaagtttaatatggagag 596
Db 534 GTTCTGT 593
Qy 597 acagtgactaccccccttgaatttgaactcgaacatgtgaatgaatgaatgaatgaatga 656
Db 594 ACGGTGACGACCCCTTGTAGTTGACTGTGACCTTGACACTTGACACTTGACACTTGAC 653
Qy 657 cataatttgcagttcagttgaagaagaacgttgcgaacttctgtcctctgaatg 716
Db 654 CATTAATTTGTCACTTCAAGTTAAGAGGAGCTTGGCAGACTTGTGTGTGTGTGTGTGT 713
Qy 717 ccaactcgaatgttgatggccatcagaagggagaccttaattctgaattatctgct 776
Db 714 CGACACTTCGATTTGGATGGCATGAGAGGGAGCTTTAATCTGAGATTATCTGCGCT 773
Qy 777 gtttaaggcaatcttctcagacttgcagtcgacggtctcaccctcagagacctatc 836
Db 774 GTTAAAGCAGTATTATTTCAGACTGAGACCGGCTCTCATCCGATGAGAGCCCTATATC 833
Qy 837 cttacgttgaagaatttgcagaagaatcctccgcaatgggttaaacatgtgataataa 896
Db 834 CTTACGTGGGAAGATTGTGAGAGATCTCCGCAATGGGTTAAACCATGCTGATATAG 893
Qy 897 ccaagaagccaggtcccccgaatccgtgctcttggagaaataaataaataacccgagca 956
Db 894 CCAAGAAAGCAGAGTCCCGAATTCGTCTTGGAGAGAAACAAACACTCGCTGTA 953
Qy 957 aaagtgcagacctctctcgtatctaacccgagatcagggagcgcgttggcgga 1016
Db 954 AAAGTCMAAGCCCTC--TCCATATATCTACCCGAGATTGAGAGCCACCGGCTTGGCCGA 1012
Qy 1017 accccaactgttcccccaccccttctacagcaagagtgctgtgagggga--cctctgc 1075
Db 1013 ACCCCATCTGTTCCTCCACCCCTTATCTGACAGGGGTGCGGAGGGAGCCTTGTGC 1072
Qy 1076 cccctcgtgagctcgtgtgtgagagacgtctgcggagactcggagccgagagagc 1135
Db 1073 CCTCTGTGAGCTCGGGGGGTGAGGAGACTGTCTGAGGAGCTCGAGCCGAGGGGGCGC 1132

Qy 1136 caccgcgagcggacagacagatcgcgatactacgctgcgacctaattggccctccat 1195
Db 1133 CACCCGGAGGGAGACACAGATCGGACATTAACCTCTCGACGTAACGGCCCTCCAC 1192
Qy 1196 gccagggggccaatttgcagccccctcagatattggcccccttctctcgcagatctataa 1255
Db 1193 ACCGGGGGGCCAAATTGACGCCCTCCAGTATTGGCCCTTTCTTGTGAGATCTGATATA 1252
Qy 1256 ttggaactaactaacatcccccttctcggagagatcccccaacgcctcaggggttggatga 1315
Db 1253 TTGGAAACTAATACCATCCCTTCTCGAGAGATCCCAACGGCTCACGGGGTGTGTGA 1312
Qy 1316 gtccctatgtctctcaccagcctaacttggatgtatttgcacagctctcgaagcact 1375
Db 1313 GTCCCTATATGTCTCTACACAGGCTACTTGGATGATGATGATGATGATGATGATGAT 1372
Qy 1376 ctccaacaccgagagcagagagaaatctgttagaggttagaataatgttccctgggac 1435
Db 1373 CTTCACACCGAGAGAGCAGAGATTTCTATTAGAGGCTAGAAAAAATGTTCCCTGGGGC 1432
Qy 1436 cgaacggcgacccacgagttgcaaaatagaattgacatggatctccctgactcgcgc 1495
Db 1433 CGACGGCGACCCACGCGGTGCAAAATGAGATTGACATGGGATTTCCCTTAACCGGCC 1492
Qy 1496 cgglttggactacaacacgctgaagtagagagacttgaatctatcgcagagctct 1555
Db 1493 CGGTTGGACTACAAACAGCGGTGAAGGTAGGAGAGCTTGAANAATCTATCGCAAGCTCT 1552
Qy 1556 ggttgagggtctccggggcgccctcagaagcgcgaactaatttggctaaagtgaagagt 1615
Db 1553 GGGGGGGGTCTCCGGGGGCGCTCAAGAGCGCCCACTAATTTGGCTAAGGTAAAGAAAT 1612
Qy 1616 gatcagagacccgaacgaacccctccgttatcttctgaagagctcagagaccttaag 1675
Db 1613 GATGACGAGGACGAGATGAACCCCTCTGTCTTTTGTGAGAGCTCTTGAAGACCTTACG 1672
Qy 1676 gcggttcaaccttctgaactcactcagagggccagaaagcctcagtgccctggcctt 1735
Db 1673 GCGGTACACCCCTTGTGATCCCACTCAGAGGCCCAAAAGCCCTCAGTGCGCTTGGCCTT 1732
Qy 1736 catlgygcagtcgctcgttgatataagaagaatactcagagacttgaagaggttacaaga 1795
Db 1733 TATAGACAGTACGCTTGTGATTAAGAAAGAGCTTGAAGAGCTGGAAGGGTTACAGGA 1792
Qy 1796 ggtcgtatcagtgatctcagtgagagcagagaagtgatctacagaagaaggagaca 1855
Db 1793 GGTGAGTTACGTGATTAAGAGAGGACGAGAGAAATATATTACAAAGGAGACAGA 1852
Qy 1856 agaggaagaagaacagagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1915
Db 1853 AGAAGAAAGGAGAAACAAAGAAAGAGAGAAAGAGAGAAAGGAGAAAGAGACTATAA 1912
Qy 1916 acggcaagaagaatttgaactaatcttggccgagtggttgaaggaagaagcagag 1975
Db 1913 ACGCGAAGAGAAATTTGACTTAAGATCTTGGCTGACGTGCTGTAAGGAGAAACCAATAC 1972
Qy 1976 ggaagagagagagatttgaagaataatgaagtcagggccctagacagtcagaggaacctggg 2035
Db 1973 GGAAGAGAGAGAGATTTAGGAATAATTAGTCAAGGCCCTTAGACAGTCAAGGAACTGGG 2032
Qy 2036 caataagaccccatctcagaagaagcagtgctglatlgtlaaagaanaagacacttggc 2095
Db 2033 CAATAGAGCCCACTCAAGAGCAATGTGATTTGTAAGAAAGAGACACTGCGC 2092
Qy 2096 aaagaaactgcacccaagaagaagaagaagaagaagaagaagaagaagaagaagaataa 2155
Db 2093 AAGGAATGCGCCCAAGAGAGAAACAAAGAGCAAGAGATCTAGCTTGAAGAAATATA 2152
Qy 2156 agattagagagagaggttctcgaacccctcccgagcccaaggttaacttgaagtgga 2215
Db 2153 AGATTAGGGAGAGAGGGGTTGCGAGCCCTCCCGAGCCAGGTAACCTTTGAAGGTGA 2212

4373 ATTAGAGCCGTACATTTTACAAAAAGCTAGCTTTTATACACTGTCTGGACATTCAGAA 4432
4436 agccaagaatcatalactagaaggaaccagatggctgaccgggttgccaagcagcagc 4495
4433 AGCTAAGATCTCATATCCAGAGAAACCATGCTGACGGGTGGCCAGCAGCGAGC 4492
4496 ccaagctgttaacctcttgccctataatagaacgcgaagcccaagcccaagcagaca 4555
4493 CCAGGCTGTATTACCTTGTGCTTATATAGAAATGCCCCAAGCCCCAGAACCCACAGCA 4552
4556 gtacaccctagaagactgccaagagataaaaga tagaaccagttctctgagactccgga 4615
4553 GTACACCCCTAAGAAAGATGGCAAGATATAAAGATAGACAGTTCTCTGAGACTCCGGA 4612
4616 ggggacctgctatacctcaatgaggaagaaatcctgcgcccaagaagaaggttagata 4675
4613 AGGGACCTCTATACCTCAGATGGGAAGAAATCTGTGCCCAAAAGAGGGTTAGATA 4672
4676 tctccaagatatacctgcttaaccaccctagaactaaacaacctgcagcagttgctag 4735
4673 TGTCCACAGATTCATCGCTTAACCCACTAGAACTAAACCTGCAGACAGTTGTGTAG 4732
4736 aacatcccttatactgtctcagagctaccaggaatgctgactcgtgtgtcaaatctg 4795
4733 AACATCCCTTATCATGTCTCTGAGGCTTACAGAGATGGCTGACTGCTGCTCAAAACAT 4792
4796 tctgacctgccaagctggttaataatgtaactctccagaactcctcaggaagaagataag 4855
4793 TGTGCTTGCACCTGCTTATATGCTTATCTTCCAGAAATGCTCCAGGGAAGAGCTAAG 4852
4856 ggggaagccaccagcgcctcaactgaggaatgtaactcactgaagtaagcggcgttaata 4915
4853 GGGGAAGCCACCCAGCGCTCAGCTGAGAACTGAGCTCACTGAGTAAAGCCGGCTAATA 4912
4916 cgggaataatactatctgtgtttgttgaacaccttcaaggaatggttagaagctatcc 4975
4913 CGGAACAAATACCTATGTGTTTGTATGACACCTTTTGAAGATGGGTAGAGCTTATCC 4972
4976 tactaagaagaagactcaacccgtgtgtgctaaagaatactcgtgagaaatcttcccaag 5035
4973 TACTAAGAAAGAGCTTCAACCGTGTGCTTAAATAAATACTGGAAGAAATTTTCCAG 5032
5036 atttggaatacctaagtaataaggtcagaacaaatggttccagcttccgttggcccaagtaag 5095
5033 ATTGGAAATACCTAAGGTATAGGGTACAGCAATGGTCCAGCTTTGTTGCTCCAGGTAAG 5092
5096 tcaaggaactggccaagatatgtggggaatgtaatggaactgcaatgtgcatatcagaccga 5155
5093 TCAAGGACTGGCCAAAGTATTTGGGGAATGATTGGAAACTGCAATGTGCAATACAGACCCA 5152
5156 aagctcagaagaatgtagaagaatgtaataacatlaagaagacccttacaataatgac 5215
5153 AAGCTCAGAGACAGGTAGAGGATGAAATGAATGAACATTAAGAACCCCTTACTAATGAC 5212
5216 cagaagaactggcaataatgtaagatggtcctccgacctgtgtctttagaggtgag 5275
5213 CGCGAGACTGGCTTATGATTTGGATAGCTCTCTGCTGCTTGTGTCTTTTAGGGTTAG 5272
5276 gaacaccccttgagacgttttggtcgaaccctctatgaattgctctacgggggaacccccc 5335
5273 GAACACCCCTTGAGACGTTTGGGCTGACCCCTATGAATTACTCTACGGGGGACCCCCCC 5332
5336 gttggaagaagaatgtccttgacataagtgctgagtgtgctgtcttcccaagcttgtctc 5395
5333 ATTGGTAGAAATTTCTTCTTACATAGTGTGACGTGCTGCTTGTCCAGGCTTTGTTCTC 5392
5396 taggtcagaagcgtcgtgaggtgtgagcagcagcagcgtggaagcagctccggagagccta 5455
5393 TAGGCTCAAGGCACTTGAGTGGGTGAGACAAAGACGCTGGAAGCAACTCCGGAGGCTTA 5452
5456 ct 5512
5453 CTCAGAGAGAGAGACTTGGAGATCCACATGCTTTCCAGTGGAGATTCAGTACGT 5512

5513 tagaagccaacgtgcaagaaacctcgaagactcgggtggaaggaacctatctctactt 5572
5513 TAGAGCCACCTGAGAGAAACCTGAGACTCGGTGGAAGGGCCCTTATCTCGACTTTT 5572
5573 gaccacacaaagcgtgtgaagtcgaagaatcccaacctgatacgtatcccaagcgt 5632
5573 GACCACACCAAGGCTGTGAAGATGGAAGATCTCCACCTGGAATCATCATCCACGT 5632
5633 taagccggggaacctcccgactcgggtggaagccgaagaactgagatcccttaa 5692
5633 TAAACCGGCGCACTCCCGATTGCGGTGGAAGAACCCGAAGACTGAATAATCCCTTAA 5692
5693 gcttcgctcacaatcgtgtgtcttactacttaacaataactcccaagcagtagtaaac 5752
5693 GCTGCGCTCCATCGCGGTGCTTCTTACTGTGTAATTAATCTCAGACTAATAGTATGC 5752
5753 gacctatagaagcgtcgaacccccctaaaccttatacctcactcgtgctgatactgac 5812
5753 GCATAGAGACAGCTGAACTCCATTAACCTTATCTCTACCTGCTTATTAATCTGACT 5812
5813 ctgatacgggtgtcaactgttaaatagcactcgaagtgctgtcctcctagaagcactgtgac 5872
5813 CCGGACAGGTATATATATCAACACACCAAGGGAGGCTCTTATAGAACCTGTGCGC 5872
5873 ctgaactgcaatctgctcctcgaattgataaccccgctgttaagaagcactcccaacc 5932
5873 CTGATCTATACGTTTGTGCTCAGATCAGTATTCCTAGTGA--CTCACCCCCAGATA 5929
5933 tagtccgtatgtatggtgtctatctgtgcagagcagcagc---agaagaagaatactgtg 5989
5930 TCTCTCATGCTCAGGATTTTATGTTTGTGCGCAGACACCAAAATATGAAATCATTTGCG 5989
5990 ggggtctcgggaatcctctcgtgaagatgagagctgcgtcaactcccaagatgagact 6049
5990 GAAATCCAGAGATTTCTTTTGAACATGGAATGGAATGCTATACCTATATGATGATTT 6049
6050 ggaatggccagatcctccagagccggtaaaatctccttctca----- 6098
6050 GGAATGGCCAACTCTCAGCAGAGATAGGCTAAGTTTCTTATTCACACCTATACCA 6109
6099 -ttccggccggcgaagtacaaatgataaactatataagaataagcgtgctcccat 6157
6110 GCTCTGACAAATTTATTTATCTGACCTGATTAAGAACTGGAAGCCCAAGTCTCTCTT 6169
6158 cagaactagatatcctaagaatgaattcact--gaagaagaagaagaataatcaca 6215
6170 CAGACCTAGATTAACCTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6229
6216 atgagataaattgtagagctggaagatagttttatataatgagcggg---ggaagcag 6271
6230 AATGGATTAATGATAGTGTGGAATGATTAATTAATTAATTAATTAATTAATTAATTA 6289
6272 gttcacttaacatcgtccttagatagagaagcggggaacaaacccctgtggcaagc 6331
6290 GCTTCATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6346
6332 gaccggaataagtaagctgctgaagcgggccccggcctgagcaccgcaataactgac 6391
6347 GACCAAAATAGGCTTTGACGGGTCAAGAACCCCAACCC----- 6385
6392 cgtgtcccaataaactcgtcgtgagcctgacataaacaagcgccttagcaagtagaca 6451
6386 -----AAGGACCAAGGACCAATCTCTATCAATTAATCTT 6415
6452 ctgattgattcctaacacaagccttagaaccctcccaagtgcttctgttgaagaagc 6511
6416 CTGATAGAGACCCCACTGATCTTAACAGCAGC-----TAAATGAGGCG 6460
6512 agaaacctcagctatcaggaagccttcaagccatcaactccacgaacgctgag 6571
6461 CAAACTTTTATGCTTATCAGAGAGCTTTTAAGCTTAACTCCAGACTCCAGAGG 6520

QY	1980	agagagagagatttaggaaaatagtcagccctagacagtcagggaccttggcaat	2039
Db	1990	AGAGAGAGAGATTTAGGAAAATTAGTCAGGCCCTTAGACAGTCAGGAACTGGGCAT	2049
QY	2040	aggaccccaactcgcacaaagaccagtcgtcgatttgaagaaaaagacacttggcaag	2099
Db	2050	AGGACCCCACTCGACAAAGGACAGTGTGATTGTAAGAAAAAGACACTGGGCAAG	2109
QY	2100	aactcccccaagaagaagaaagaccgaaggtccttagctctagaagaagataaagat	2159
Db	2110	AACCTCCCCAAGAAAGGAAACAAAGACGAAAGTCTTAGCTCTAAGAAGAAATAAAT	2169
QY	2160	taaggagagcgggttcgagcccccctcccgagcccaaggttaacttgaagtggaagg	2219
Db	2170	TAGGGGAGACGGGGTTTCGGACCCCTCCCGAGCCACAGGTAACCTTTGAAGTGGAGCG	2229
QY	2220	caaccagttgattccctggttgataccggagcggagcaatcagtcgtctacaacatla	2279
Db	2230	CAACCAGTTGAGTTCTGTGTGATACCGAGCGGAGCATTCAGTCTCTACAAACATTA	2289
QY	2280	ggaataactaaagaataaaatccctgggtgaatgggtgcacagggcaacggcaatla	2339
Db	2290	GGAAACTAAAGAAAGAAAAATCCGGGTGATGGGTGCCACAGGGCAACGGCACTATCCA	2349
QY	2340	tggactaccggaagaccgttgaacttggagtggaacgggttaacccactcgttctggtc	2399
Db	2350	TGGACTACCCGGAAGAACGGTTGACTTGGAGTGGGACGGGTAAACCACTCTTTCTGTGTC	2409
QY	2400	atccctgagtcgccagtaaccctctcagtagagagacttaactgaccagaatggagactoa	2459
Db	2410	ATCCCTGAGTGCCTCCCTTCTAGTGTAGAGACTTACTGACCAAGATGGAGCTCAAA	2469
QY	2460	attctcttgacaagaagaccgaaggtgctcttgaaatacaaacatcactgctgttg	2519
Db	2470	ATTCTTTTGAACAAAGAAAGACCAAGAGTGTCTGTGAATTAACAACCATACTGTGTG	2529
QY	2520	accctcaatlagatgataatcogataatctccccaagtaaaagcctgatacaagat	2579
Db	2530	ACCCTCAATTAGATGATGATATCGACTATTTCTCCCAAGTAAGCCGTGATCAAGAT	2589
QY	2580	atacgtctcgttggagacattccccaagccttgggcagagaaccgcagggatgggttg	2639
Db	2590	ATACGTCCTGTGGAGACATTTTCCCAAGCTGGGCAGAAACCGGAGGATGGTTTG	2649
QY	2640	gcaagcaagttcccccacaggtatcaactgaagccagtgctacacagatcagtc	2699
Db	2650	GCAAGCAAGTTCCTCCCAAGTTATTCATCTGAAGGCCAGTGTACCAAGTATCAGTC	2709
QY	2700	agacagtaacccttgaatagagagcctcgaagaaatttggccgcagtgtcaagaata	2759
Db	2710	AGACAGTACCCCTTGAGTAGAGAGGCTCGAAGAAATTTGGCCGCAATGTTCAAAAGATTA	2769
QY	2760	atccaacaggacatccctagctcgtcccaatcccttggaaactcctccctgctacaggtt	2819
Db	2770	ATCCAACAGGGGATCTAGTTCCTGTCCTCAATCCCTTGGAATATCTCCCTGTAACCGTT	2829
QY	2820	aggaagccttggaccaatgattatcgaacagtagcaggacttgaagaggtgcaataaag	2879
Db	2830	AGGAAGCCTGGACCAATGATTATCGACAGTACAGGACTTGAGAGAGGTCAATTAAGAAG	2889
QY	2880	gtcgaagacatatacccaaggtcccgaaacctataaccttgaagcgcctccgcct	2939
Db	2890	GTGCAAGACATACACCAACAGGTCCCAACCCCTTATTAACCTTGAGCGCCCTCCGCT	2949
QY	2940	gaacggaacttggtaacagtaattgagcttaaaagatccttctctcgcagatlaaac	2999
Db	2950	GAAAGGAGTGTGTACAGTATTTGACTTAAAGATGCTTTCTCTGAGATTACAC	3009
QY	3000	cccaactagccaacacttctgctctcgaatgtagagagatccaagttaaggaagaccgg	3059
Db	3010	CCCACTAGCCAAACCTTTTACTCTGTAATGAGAGATCCAGGTACGGGAGAAACCGGG	3069

QY	3060	cagtcaacttggaccgcacttgcaccaagggttcaagaactcccccagaccatttgaagaa	3119
Db	3070	CAGTTTACCTGTGACCCCACTGCCCCAAGGGTTCAAGAACTCCCGACATCTTTGACGAA	3129
QY	3120	gacctacaagggaccttgcacaaacttcaagatccaacacccctaagtacccctccacag	3179
Db	3130	GCCCTACACAGGAGCACTGGCCAACTTCAAGATCAACACACCTCAAGGTGACCTCTCCAG	3189
QY	3180	taagtgaatgacctgctctcgtcgtggagagcccaacaagagacttcttagaagtagaag	3239
Db	3190	TACGTGATGACCTGCTTGTGGGGAGCCACCAAAACAGGACGTGTAAGAGGTACGAAG	3249
QY	3240	gcaactgcttggaaattgtctgaacctagagctatagaagccctcttagaagggccagatt	3299
Db	3250	GCACTTCTGCTGCAATTTGTCTGACCTAGGCTACAGAACCTCTCTTAAGAAAGCCAGATT	3309
QY	3300	tgcagagagaagttacaatacttgggttacagtttgcggggcgggagcgaatgctgagc	3359
Db	3310	TGCAGAGAGAGAAATCACTTGGGGTACAGTTTGGGGGGGGGAGCGATGCTGACG	3369
QY	3360	gaagcaaggaagaanaactgtagtccagataccggcccccaacacgaacgaagtga	3419
Db	3370	GAGGCACGGAAGAAACTGTAGTCAGATVACGGCCCAACACACGCAAAAGTGAGA	3429
QY	3420	gaatttttgggaagcttgatttgcagactgtgatacccggtgttgcagcttagca	3479
Db	3430	GAGTTTGGGGACAGCTGATTTTGCAGACTGTGATCCGSGGTTTGCAGACTTAGCA	3489
QY	3480	gccccactctaccgcctaaccaaagaagaaggagatctccttggctcctgagcaag	3539
Db	3490	GCCCCACTCTACCCGGTAAACCAAGAA-----	3516
QY	3540	aaggcaattgtagtctalcaaaaagccctgtcgtagcagcaactgtcttgcctccctgac	3599
Db	3517	-----	3516
QY	3600	gtactaaacccttlaaccttlatgtlgaatgagcgtlaaggagtagccgagagattla	3659
Db	3517	-----	3516
QY	3660	accocaaacctaggaacatgagagagacctgttgcctacttgcacaaagaagcttgact	3719
Db	3517	-----	3516
QY	3720	gtagcagtggttggcccgatgtctgaaggtcatcgcagctgttgcatactggtcaag	3779
Db	3517	-----	3516
QY	3780	gaagctgacaatltgacttggagacgaalatalactgtaatagcccccatgttgag	3839
Db	3517	-----	3516
QY	3840	aacatggttcgcagcccccagacgcatgtagtgaaccaagcccgcatgaaccaatla	3899
Db	3517	-----	3516
QY	3900	agcctgtcttcaacagagaggtcacttctgcctcaacacagccgcttcaacctgcaact	3959
Db	3517	-----	3516
QY	3960	cttctgcttgaagagactgtagaacagtgactcatgatttgcataactatltgag	4019
Db	3517	-----	3516
QY	4020	gagactggggttccgcaaggaaccttaagacataacgcgtcgtgaggaagtgtctaacttgg	4079
Db	3517	-----	3516
QY	4080	ttcaactgaggaagcagctatgtgtgtgaagttaagagtagtctggggcggcagltgtg	4139
Db	3517	-----	3516
QY	4140	gacgggaccgcagcatcttggccagcagccttgcgggaaggaacttcaagcgcaaaagct	4199

Db 3517 ----- 3516
Qy 4200 gagctcatgcccctcaagcagccttgcgctgcggaaggaaatccataacattat 4259
Db 3517 ----- 3516
Qy 4260 aagcaagcaggtatgcttgcagctgacagctacagggccatctataacaag 4319
Db 3517 ----- 3516
Qy 4320 gggcttctacccagcaggggaaataaagaacaagaagaattcttaagctta 4379
Db 3522 ggggtgcttaccctgacgaggggaaataaagaacaagaagaattcttaagctta 3581
Qy 4380 gaagccttaattgcccagaagaagctgctatatacaactgctctgacatcaagaagc 4439
Db 3582 GAAGCCTTACATTTGGCCAAAAGGCTACCTATATACACTGTCTGGACATCAAGAAGCC 3641
Qy 4440 aaagctcatatctagaagggaaccagaatgctgacgggtctgccaagcagcagccag 4499
Db 3642 AAAGATCTCATATCTAGAGGGAACCAAGTGCACCGGTTGCCAAGCAGAGCCAG 3701
Qy 4500 gctgttaacctctgctataatagaacagcccaagcccaagaccagaagcagctac 4559
Db 3702 GCTGTAACTCTTGCCCTATATAGAAAGCCCAAGGCCCAAGCCAGACAGATAC 3761
Qy 4560 accctagaagactgcgaagagataaagaatagaacagctctctgagactccggaagg 4619
Db 3762 ACCCTAGAAGACTGGCAAGATATAAAGATAGACAGTCTGTGAGACTCCGAGGG 3821
Qy 4620 acctgctacatctatggaagaaatctctgcccagaagaagggttaagatctc 4679
Db 3822 ACCTGCTTACTCTATAGGGAAGAAATCTGCCCCAAGAAAGGTTAAGATATGTC 3881
Qy 4680 caacagatatacgtctcaaccacctaagaactaaacactgcagcagttgctagaaca 4739
Db 3882 CAACGATACATCGCTCAACCCACCTAGAGACTAAACACTGCAGCAGCTTGTCAGAAA 3941
Qy 4740 tcccttatactgtctgagctccaggaatgctgactgctgctgctgctgctgctgctg 4799
Db 3942 TCCCTTATCATGTTCTGAGGCTACCCAGAGATGCTGACTGCTGGGCTTAAACATTTG 4001
Qy 4800 cccctgacagctggttaagctcaatccctcagaatataccctcagaagaagaactaagg 4859
Db 4002 CCGTCCAGCTGTTAATGCTAATCCTTCCAGAAATACCTCCAGAAAGAGACTAAGGGA 4061
Qy 4860 agcaaccagagcgtcactggaagtggaactcactgagtggaagccgagctaaatcaga 4919
Db 4062 AGCCACCCAGGCGCTCACTGGGAAGTGACTTCACTGAGTAAAGCCGGCTAAATAGGA 4121
Qy 4920 aacaaatctatgtgttggtagaacctctcagaatgaggtgagagctactact 4979
Db 4122 AACAAATATCTATGTTGGTTTGTGACACCTTTTCAGAGATGGGTAGAGGCTTATCT 4181
Qy 4980 aagaagaagactcaaccgtggtgctagaanaaatactggaagaattttccaagaatt 5039
Db 4182 AAAAAAGAGACTCAACCGCTGTGGCTAAGAAATACTGAGGAATTTTCCAAAGATT 4241
Qy 5040 ggaatacctaaggtataggttcagacaaatgctcagcttgcgttgcagagtgatcag 5099
Db 4242 GGAATACCTAAGTAAATGGGGTCAGACAAATGCTCAGCTTGGTGGCCAGAGTAAAGTCAG 4301
Qy 5100 ggaatgccaagatactggagatgagtggaactgcatctgctatagaccccaagc 5159
Db 4302 GGACTGGCCAGATATTTGGGGATTGATTGAATACTGCTTGTGATACAGACCCCAAGC 4361
Qy 5160 tcaagacaggtagagagatgaatagaacaaatlaagaagaccccttaacaaatgaacaa 5219
Db 4362 TCAGAGACAGGTAGAGAGATGAATGAACATTAAGAGACCCCTTACCAATTTGACCA 4421
Qy 5220 gagactgcatatgattgagctgctcctgctgcttgccttttaagggttaaggagc 5279
|||||

Db 4422 GAGACTGGCATTAATGATTTGGATGAGTGCCTCCCTGCTTTGCTTTTAAAGGTAGGAAC 4481
Qy 5280 accctggaagatcttgagctgaaccctataagattgctctacaggggaaccccccgctg 5339
Db 4482 ACCCTGGACAGTTTGGGCTGACCCCTTAAATTTCTTACGGGGAGACCCCGCTTG 4541
Qy 5340 gcagaattgctcttgcaatagtgctgaatgctgctcttccagcttctgctcag 5399
Db 4542 GCAGAAATTTGCTTTCACATAGTGTGATGCTGCTTTCCAGCCTTGTCTTCAAG 4601
Qy 5400 ctcaagcgtcgtgagtgagtgagcagcagcagtggaagcaatcccggaagcctactca 5459
Db 4602 CTCAAGCGCTGCAAGTGGGTGAGGAGCGAGCGTGAACACCTCCGGAGGCTTACTCA 4661
Qy 5460 ggaaggaacttgcaagttcacatctgctcccaagtggagatcagatctagtcagc 5519
Db 4662 GGAGGAGACTTGCAGTTCACATGCTTCCCAAGTTGGAGATTCACTATTTAGAGCC 4721
Qy 5520 caccgtgcaagaaacctcgagactcgglygaagggaacctatctgactttgaacaca 5579
Db 4722 CACCGTGCAAGAAACCTCGAGACTCGTGAAGGAGACTTATCTGACTTTGACCA 4781
Qy 5580 ccaagcgtcgtgagtgagtgaggaatcccaacttgatccatccagcttaagc 5639
Db 4782 CCAACGCTGTGAAGTCAAG----- 4803
Qy 5640 ggcacacctccgactcgggtgagagcgaagaagctgagatcccttaagctcgc 5699
Db 4804 ----- 4822
Qy 5700 ctccatgcctgcttctactctataaactcccaagccagcagtgatgaagcctat 5759
Db 4823 CTCATGCGCTGCTGCTTCTTACTCTGCAATTAACCTCTCAAGTTAAGTAAAGCGCTTGT 4882
Qy 5760 agacagctggaacccccatagactcttacccttactgctgctgctatgagcccgatag 5819
Db 4883 GGACAGCGCCCAACTCCATTAACCTTATCTCTACCTGCTGTTACTGACTCCGGTAC 4942
Qy 5820 gggctgcaactgtaataagcactcagagtgctgctcctagaagcactgctgagcct 5879
Db 4943 AGGTATTAATTAATACAGCACTCAAGGAGAGCTGCTTGGGAGCTGCTGAGCTGAAT 5002
Qy 5880 gcaatctgctcagatgatacccgctgta-----aaagcaacctcccaact 5933
Db 5003 ATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5062
Qy 5934 agtcgagtgatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 5990
Db 5063 ACTCGGCTTACGGGTTTACCTTTGCCCCAGAGACCCCAATTAATGAAGAAATATTG 5122
Qy 5991 gggctcgggaatcctctctgtagagatgagtgagtgagtgagtgagtgagtgagtgag 6050
Db 5123 AAATCGTCAGAGATTCTTTTGAAGCAATGAGACTGCTAATTAATGATGAGCAATG 5182
Qy 6051 gaaatgcccagatctctccaggaacccgggttaaatctcctctgtaactcc----- 6102
Db 5183 GAATGCGCAGCTCTCAGCAAGACAGAGAAATGCTTCTTTGTTAAACATCTTACAG 5242
Qy 6103 -----99ccgggcaagtaacaaa 6121
Db 5243 TTTAATCAATTTAATTTATGCGCATGGAGATGGAAAGATTTGCAACACCGGTTACAAA 5302
Qy 6122 tgaataactatataaagataagcgtgctcccatcagactagatatactaaagataa 6181
Db 5303 AGATG-TAGCAAAATGAAGCAATTAAGCTGATTCGTTAGCCATGATTTCTTAAATA 5361
Qy 6182 gttcactg--aaagaaacaggaataatctaaagatgagataatgtagagctgag 6239
Db 5362 GTTCACTGAAAAAGAAAAAAGAAAAATTTCAAAAAGGGGTAATGATTAATCTTGGG 5421
Qy 6240 gaatagtttttat-----tatagcgggggagcagagtgccatcttaacatcgcccta 6295
Db 5422 GAATAGTACTATAGAGGCTTGGGAGAAAGAAAGAGATCTGTTCTGACTATTCCGCTCA 5481
|||||

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-766-528-1

Query Match 60.7%; Score 4986; DB 4; Length 8060;
Best Local Similarity 95.28; Pred. No. 0;
Matches 5276; Conservative 0; Mismatches 245; Indels 23; Gaps 12;

QY 1 gtgtgtacagacttgagcccaagcgccttgagataaaatcccttgctgttgcac 60
DB 2530 GTGGGTGACGACTGTGGGCCCCAGCGCTTGGAATAAATCCTCTTGCTTGATC 2589
QY 61 aagacgcctctcgtgagtgat--tgggtgtgcctctccagcccgagaggggga 119
DB 2590 AAGACCGCTCTCGTAGTGAATTAAGGAGATCCCTTTTCCGAGCTGGAGG----- 2642
QY 120 ttgtcttctacgtccttcatattgtgtgctgagcgaggaatcccgagacacccct 179
DB 2643 --TTCTTTTCTGCTTACATTTGGGGCTGCTCGGAGTCTGTGCGGCCACCCCTTA 2700
QY 180 acaccgagaaacgacttgagagta-aagagatccctcttgaaacgtgtg--tgtatc 236
DB 2701 ACACCCGAGAACCGACTTGAGGTAATAAAGATCCTCTTTTAACGATGATGATGACC 2760
QY 237 ggcgcgctctctgttcgagtgctgttctcgatgagcgcgcttccggtttgcaatg 296
DB 2761 GGCCTGCTCTGTCTGAGTGTCTGTTTCAGTGTGCGCGCTTTCGGTTGCACCTG 2820
QY 297 tccctcagacgtaagagcagactgtatcagcaagacgtctagagagatcaacag 356
DB 2821 TCTCTCAGGCGCTAAGGCGTGGGGGACTGTGATCAGACGAGCGTAAAGAGATCAAG 2880
QY 357 gctgcaccccttgaggagccccgggaggttgaggagagcaaggagcgctgtgtctcc 416
DB 2881 GCTGCTGCTCGGGGAGCCCGGAGAGTGAAGAGCCAGGAGCGCTGTGTGTCTCC 2940
QY 417 tacttgagtagaagagcagactctgttcttgaaacgaagacttcccccgcgcgcg 476
DB 2941 TACTTGGGTAGAGAGCCGAATTTCTGTTCTGAAGCCAAAGCTTCCCCCTCCGAGACG 3000
QY 477 tccgactcttgcctgtcttgaaagacgcgagcggttcgctgtgtcttgatctgtg 536
DB 3001 TCCGACTCTTTTGCTCTGTGGAATACTGTGAAGGGGTACAGTGTCTGTGATCTGTG 3060
QY 537 gtctctgtcgtgtgtcttctgtctgtgtgcttctgtctacagtttaataatggagag 596
DB 3061 GTTCTGTTTGTGTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
QY 597 acagtgatcaccccccctagtttgactcagacattgagactgagtaagttagccagggt 656

DB 3121 ACGGTAGACACCCCTCTAGTTGACTCTGACACTTGAGAGTGAATTAATCCAGGCT 3180
QY 657 cataattgtcagtttaagtttaagagagacttggcagacttctgtcccttgatgg 716
DB 3181 CATTAATTTGTCAAGTTCAAGTGAAGAGGACCTTGGCAACATTTCTGTCTGAATGG 3240
QY 717 ccaacattgattgttagtgcacatcagaagggagaccttaattcgaatlaacctgct 776
DB 3241 CCGACATTTGATGTTGATGAGCCATCAGAGGGACCTTTAATTCTGAGATTACTGTGCT 3300
QY 777 gtttaagcaatcatttctagaactgagccggtctctatcctgatcagaagccctatc 836
DB 3301 GTTAAAGCAATTAATTTTTCAGACTGAGACCCGCTCTCAATCCGATCAGAGGCCATATATC 3360
QY 837 cttagtgcaagatttgacagaagaatctccgcacatgggttaaacatgtgtaataaa 896
DB 3361 CTTACGTGCAAGATTGTGCAAGAGATCTCGCCATGGGTTAAACATGCTGTAATTAG 3420
QY 897 ccaagaaagccaggtccccaatcctgtgctcttgagagaaacaaacactcgcgaa 956
DB 3421 CCAAGAAACCAAGTCCCGCAATTTGCTTGGAGAGAAAACAAACACTCGGCTGAA 3480
QY 957 aagtcgagccctcttctcgtatctaccgcgagatcgagagcgccgacttgccgga 1016
DB 3481 AAGTCAAGCCCTTC--TCCATATATTAACCCGAGATGAGGAACCAACGCTTGGCTGGA 3539
QY 1017 acccaacactgttcccccaccccccctatccagacagaggtgtgtgagggga--cctctgc 1075
DB 3540 ACCCCATTTGTTTCCCCACCCCTTATCTGACAGAGGTGCGGAGGGAGACCTTTGTC 3599
QY 1076 cctctctgagctcgtgtgtgagagacgtctgcggagactcggagccggagagcg 1135
DB 3600 CCTCTCTGAGCTCCCGCGGCTGAGGAGACCTTCTGACGAGACTCGAGCGGAGGGCGGC 3659
QY 1136 caaccgagagcgagacagagagatcgagattacgcgtgcgacctaagggccctccat 1195
DB 3660 CACCCCGAGCGGACAGAGAGATCGCATTAACGCTGAGCCAGCTACGCGCTCCAC 3719
QY 1196 gccaggggccaatttgaaacccctccagatattggcccttctctcagatcctataa 1255
DB 3720 ACCGGGGGGCAATTTGACAGCCCTCCAGATTTAGGCCCTTTTCTGTCAGATCTATPAA 3779
QY 1256 ttggaactaacatcccccttctcggagagatcccccaagcccaaggggtgtgtga 1315
DB 3780 TTGGAATACTAACCATCCCTTTTTCGGAGATCCCCCAACCTCACGCGGTTGTGTGA 3839
QY 1316 gtccctatgtctctcacacagcctacttggatgatttgaacagctgtcagacact 1375
DB 3840 GTCCCTTATGTCTCTCACAGCTACTTGGGATGATTGTCAACAGCTGTCTCAGACACT 3899
QY 1376 ctccaacacgagagagagagaaatctgttagagctagaacaaatgttccctggggc 1435
DB 3900 CTTCACAAACCGAGAGACGAGAGAAATTTCTATTAGGCGCTAATAAAAAATTCTCGGGC 3959
QY 1436 cgaaggggagccagcagatttgcaaaatagatttgacatggagattcccttgaatcgcc 1495
DB 3960 CGAGGGCGACCCACCGGCTGGAATAAGATGACATGGGATTTCCCTTAATCTCGGCC 4019
QY 1496 cgtttgagactacaacagcttgaagtagagagacttgaacatctacgcagagctc 1555
DB 4020 CGGTTGGAGATTACAACCGGCTGAAGTGAAGGAGCTTGAATAATCTATCGCAGGCTCT 4079
QY 1556 ggtggcggtctccgggggcctcaagaacggcccaataattgtctaaagttaagaggt 1615
DB 4080 GGTGGGGGCTCTCGGGGGCCCTCAAGACGGCCCAATTTTGGCTTAAGTAAGAAAT 4139
QY 1616 ggtgacagagcagaaacacccctcgttatttcttgagagagctcatgaaagcctcag 1675
DB 4140 GATGCAAGGAGCCAAATGAACCCCTCTGTGTTTCTTGAGAGGCTTTTGAAGCCTTCAG 4199
QY 1676 gcggttcaacccttctgactcactcagaaggcccaagagcctcagtgccctggcct 1735

Db 4200 GCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAAGCCTCAGTGCGCTTTGGCCTT 4259
QY 1736 catlgygcagtcgctcgtgatatcaagaagaactcagagactcgaaggttaacaga 1795
Db 4260 TATAGACAGTCAGCCTTGATATTTAGAAAGAGCTTCAGAGACTCGAAGGGTTATACAGA 4319
QY 1796 ggcctagtgatcgtatcgtatgagagagcagaaggttattacagaagagagacaga 1855
Db 4320 GGCCTAGTTACCTGATCTAGTGAAGAGCGCAGAGAAAGTATATTACAAAAGGAGACAGA 4379
QY 1856 aagaggaaggaagcagagagaaagagaagagaagaagagagagagagagagagagag 1915
Db 4380 ACAAGAAAGGGAACAAGAAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4439
QY 1916 acggaagagaagaatttgaactaagatctggccgcaatggttgaaggaagagagagagag 1975
Db 4440 ACGGCAAGAGAGATTTTGATTAAGATCTTGCTCAGTGGTTGAGGAAAGCAATAC 4499
QY 1976 ggaagagagagagagatlttaagaaaattagtcagccctagacagatcaggaacctggg 2035
Db 4500 GGAAGAGAGAGAGATTTTATAGGAAATTTAGGTACAGGCCCTAGACAGTACAGGAACTGG 4559
QY 2036 caatagagaccccaactcgaagagagacagtgatgctatgttaagaagaagacactggc 2095
Db 4560 CAATAGAGACCCACTCGAAGAGAGACCAATGTGCATATTGTAAAGAGAGAGAGAGAGAG 4619
QY 2096 aaggaactgcccgaagagaagaaacaagaagcgaaggttctagctcagagagaagataa 2155
Db 4620 AAGGAATCGCCCAAG 4679
QY 2156 agattagagagagagaggtlctggagcccccctccgaagccagaggaacttgaagtgga 2215
Db 4680 AGATTAGGGGAGACGGGGTTCGGACCCCTCCCGAGCCCAAGGGTAATTGAAGTGGA 4739
QY 2216 ggggaacacagttgagttcgtcgttgaatcagagagagagagagagagagagagagag 2275
Db 4740 GGGGACACCACTGATGATTCCTGTTGATACCGGAGCGAATTCAGTGCATACAGCC 4799
QY 2276 attaggaactaaagaagaaataatccctggtgatggtgcacagaggaagcagagcagta 2335
Db 4800 ATTAGGAACCTAAAGATTAATAATCTGGGTGATGGTG -CACAGGGCAACACAGTA 4858
QY 2336 tccatgagactaccgaagaacggttgaaggtggaggtggaacccaactcgttct 2395
Db 4859 TCCATGAGCTACCCGAAAG -ACAGTTGACTTGGAGTGGAGCGGGTAACCCACTGTTCT 4917
QY 2396 ggtaacccctagtgcccaagtaacccttctagtgtagagactactgacaaagatggagac 2455
Db 4918 GGTCAATCTGATGTCGCCAGCACCCCTTTAGGTAGAGACTTATTGACCAAGATGGAGAC 4977
QY 2456 tcaaatctcttgaacaagaagacagaagtgctcgtgaataacaacccaactcagt 2515
Db 4978 ACAATTTCTTTGAGACAGGAGAAACCAAGAGTGTCTGCAATATACAAACCTATACCTGT 5037
QY 2516 gttgaacctccaattagatgaatatacgaataatcttccccaagtaagccgtga tca 2575
Db 5038 GTTACCCCTCAATTAGATGAGATATCGACTTACTCTCCCTAGTAAGAGCTGATCA 5097
QY 2576 agataacagtcctggttgtagagcagtttccccaagcctcggcgaagacccgaggaatggg 2635
Db 5098 AAATATACATTTCTGTTGGTAGACATTTCCCAAGCCTGGGCAAAAACCCAGGAGTGGG 5157
QY 2636 tttagcaagcaagtlccccaacaggtatcaactgaagccagctcagacagatc 2695
Db 5158 TTTGGCAAGCAAGTTCCTCCCAACAAGTTTATTCACACTGAAGCCAGTCCACACAGTGC 5217
QY 2696 agtcaagacgttaaccttgtagtagagagctcgaagaagaatttggccgcatgttcaag 2755
Db 5218 AGTAGACAGTACCCCTTGAATTAAGAGCTCAAGAGAGATTCGGCCGATGTCCAAAG 5277
QY 2756 attaatcaacagagcatcctcagttcctgtccaaatcccttgaaataatcctctgacac 2815
Db 5278 ATTATCCACAGGGAATCTTCTGTCTGATCTGCTGCAATCTCCGAAATACCTCCCTGTACC 5337

QY 2816 gtttagaagccttggaaccaatgatatcgaacagtaacaggaacttgaagaggtcaataa 2875
Db 5338 GGTTAGAAGGCTTGAGACTATATGACTATCGACCGGTACAGAGACTTGAGAGAGTCAATTA 5397
QY 2876 aaggtgcagacatacaaccaacggtccgaacccctataacctcttgagcgccctcc 2935
Db 5398 ACGGTCAGAGATATACCAACCAACGACGCGGAACCCCTTATACCTCTTGTGTCTGCC 5457
QY 2936 gctcgaacaggaactggttaacagatattggaacttaaaagatgcttctctgctagat 2995
Db 5458 ACCCAACGAGAGCTGATACAGTATGACTTAAAGATGAGCTTCTTCTGCTGAGATT 5517
QY 2996 acaccccaactagccaacacacttlttgcttcgaatgtagaagatcccaaggttcaggagaac 3055
Db 5518 ACACCCCACTACCAACCACTTTTGTCTCGATGAGAGATCCAGGTCCGACATCTTGA 5577
QY 3056 cgggcaactcaacttgaccacccgactgcccgaaggttcaagaactlcccgacacacttga 3115
Db 5578 CGGCACTCCTACCTGACCCGACTGCCCAAGGGTTCAGAGATCCCGCCGACATCTTGA 5637
QY 3116 cgaagccctcaacagaggaactcgtgccaactcagatccaaacccctcaggtgacctcct 3175
Db 5638 CGAAGCCCTACAGAGAGACCTGGCAACTTCAGAGATCCAAACACCTCAGGTGACCTCCT 5697
QY 3176 ccaatagctgtagaacctgtctctgagggagagccaccaacaggaactgcttagaagatc 3235
Db 5698 CCAATAGCTGATGACTGCTTGTGCGGAGACCCACCAACAGAGACTGTTTAAAGAGCAC 5757
QY 3236 gaagcactactcgtgnaatgtctgacctagagctcagagccctcgtcctaagaagccca 3295
Db 5758 GAAGGCACTACTCTGGAAATTTGCTGACCTAGAGCTACAGAGCCTTGCTTAAGAGCCCA 5817
QY 3296 gatttgagagagaggttaacataacttgggtacagtttgcgggctgggcaagatagct 3355
Db 5818 GATTTCAGAGAGAGAGATACATATTGGGTACAGTTTACGGAGAGGGGACGAGTGGCT 5877
QY 3356 gaaggaagcaggaagaanaactgtagtccagataccggccccaacacagcccaaaagt 3415
Db 5878 GACGAGAGCGAGAGAGAAACTGTATGCTCGATACCGGCCCAACACAGCCCAAAACAT 5937
QY 3416 gaagagtttggggacagctggaatttgcagaacttggatccgggggttggcgaact 3475
Db 5938 GAGAGATTTTGGGACACACTGATTTGACAGCTGGATCCGGGGTTGGCGACCTT 5997
QY 3476 agcagcccaactctaccgctaaacaaagaagggagattctcctgggtctcgaagca 3535
Db 5998 AGCAGCCCACTACTACCCGCTAACCAAGAAAGGGAATTCCTGGGCTCTGAGCA 6057
QY 3536 ccagaaggaatttgatgatacaaaaagccctgctgagcgaacccctgctggccctccc 3595
Db 6058 CCAAGAGGCAATTGATGCTATCAAAAAGGCGCTGAGCGCACCTGCTTGCCCTCC 6117
QY 3596 tgaagtaactaaaccccttaaccttaltgtgtagtgagcgtlaagggagtagccgaaggt 3655
Db 6118 TGACGTAACTAAACCCCTTACCTTATGTGATGAGCTGAAGGGAGTGGCCGGGGAT 6177
QY 3656 tttaacccaacccctaggaacatggaaggaactgttgcctactcgttcaaaagacttga 3715
Db 6178 TTTAACCCAAACCCCTAGGACATGAGAGAGACTGTGCCCTCACTCAAAAGACTGCA 6237
QY 3716 tccgttagccagtggttggcccgatgtctgtagaggtatccgaagcttggcaatctgt 3775
Db 6238 TCTGTAGCCAGTGGTGGCCCATATGCTGTGAAGGCTATTCGAGCTGTGGCATCTGGT 6297
QY 3776 caagagcgtlgacaaattgacttggagc -agataataactgtlaatagcccccata 3834
Db 6298 CAAGGAGCTGACAAATTTGACTTTGGACAAGAAATATACTGTAAATAGCCCCCATGAT 6357
QY 3835 tggagaaactggttggagagcccccagacaggttgaatgaacagcccgcatgacccact 3894
Db 6358 TGAGAAATATGTTGGCGAGCCCCCAAGACGATGATGACCAACGCCCGCATGACCACT 6417

QY 3895 atcaaacctgctctacagagagaggaatcttgcctccacacgacccgctcaaacctg 3954
|||||
Db 6418 ATCAAAACCTGCTTCTACAGAGAGGGGTCAAGTGGCTCCACCAACCGCTCAACCTG 6477
QY 3955 ccaactctctcgcgaagaactatgaacacgtgactcaatgttgcacatcaactatga 4014
|||||
Db 6478 CCACCTCTTCTGCTGAGAGAGACGTATGACACAGTGAATGATGCTCACTCACTATTTGA 6537
QY 4015 ttgagagactcgtggtccgcaaggaacttacagacataccgctgacgtgaggaagtctaa 4074
|||||
Db 6538 TTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACCTGGAAGTGTAA 6597
QY 4075 cctgttcaactgacgaagaacgacatgtgtgtgaggaagtaagagatgtgcgagcgag 4134
|||||
Db 6598 CCTGGTTCACCTGACGGAAGGACGTATGTGGTGGAGTAAGAGATGGCTGGGGGGGG 6657
QY 4135 tgggtgacgagaccgcaagactcgtggccagcgtgcggaagaaactcagcgcaaa 4194
|||||
Db 6658 TGGTGAACGGGACCCGACAGATCTGGGCGCAGACCTGCGGGAGGAATCTCAACAAAA 6717
QY 4195 aggtctgagctatgtccctacagaaacttgcggtgcgcaaggaagaaatccataaca 4254
|||||
Db 6718 AGGCTGACCTCATGCGCTTACGCAAGCTTGGCGGTGGCCGAAGGAAATCCATAAACA 6777
QY 4255 ttatacagacagacagatgtccttgcgactgacacgtacacggtggccatcatalaac 4314
|||||
Db 6778 TTTATACGAGACAGAGATATCCCTTTGGAGTGCACACGTACATGGGGCCATCATTAAC 6837
QY 4315 aaaggggtgtcttaacctcagcagggaggaataaagaacaaaggaatcttaagcc 4374
|||||
Db 6838 AAAGGGGTGTCTTACCTTACCTCAGGAGGAGGAAATTAAGAACAAAGAGAAATTTAAAGCC 6897
QY 4375 tattagaagcttatcttgcacaaagagcttatatacactgtccctgagacataga 4434
|||||
Db 6898 TATTAGAAGCGGTACATTTACCAAAAAGGCTAGCTATTATACACTGTCTTGACATAGAGA 6957
QY 4435 aagccaagatcatalatcagaggaacacagatgtcgtgacgggtgtgccaagcagcag 4494
|||||
Db 6958 AAGCTAAGATCTCATATCCAGAGAAACAGATGGCTGACCGGGTGGCCAAAGCAGCAG 7017
QY 4495 cccaggtcgttaacctctgcctataagaacgcccccaagcccccaagacacagac 4554
|||||
Db 7018 CCCAGGCTGTTAAACCTTGTGCTTAAATAGAAATGCCAAAGCCCAAGAACCCAGACAGC 7077
QY 4555 agtacacctgaagacgtgcgaagaataaagaatagacagcttctcgtgacgtcgg 4614
|||||
Db 7078 AGTACACCTTGAAGACACGTGCAAGAGATAAAGAAAGATAGACCA-TTCTCTAGACTCCGG 7136
QY 4615 agggagactgtctatacctcatalatggaagaatcctgcgccacaagaaggtagaat 4674
|||||
Db 7137 AAGGAGCTGCTATACCTCAATGAGGAGAAATCTCCGCCCAAGAAAGGTTAGAAAT 7196
QY 4675 atgtccaac-agatatacgtcttaacccaactaggaactaaacccgtcagcagttgctc 4733
|||||
Db 7197 ATGTCCAACAGATACATGCTTAACCCACTAGGAACCTTAACACCTGACACCAATTTGCTC 7256
QY 4734 agaacatccctcatcatgctctcgtacaggtacagagatgtgctgctgctgctaaacat 4793
|||||
Db 7257 AGAATATCCCTTATCATCTGTTCTGAGGCTACCAAGAGAGTGGCTGCTGCTAAACAT 7316
QY 4794 tgtgtccctgcagcagctgttaatgtataccttccagataactccagaaagagacta 4853
|||||
Db 7317 TGTGTGCGCTGACACTGTGTTAATGTCTAATCTTCCAGAAATGCGTCCAGGAGAGACTA 7376
QY 4854 aggggaagccacacagcgtcactcgtggaagtgtgactcactgaggttaagccggtctaa 4913
|||||
Db 7377 AGGGGAAGCCACCCAGGGGCTCAGTGGGAAGTGGACTTCACTGAGGTAAACCCGCTAAA 7436
QY 4914 taaggaaacaaatctatgtgtttttagaacacctttagaatgtgttagaggtat 4973
|||||
Db 7437 TATGAAACAAATACCTTTTGGGTTTTTTTATGACACCTTTTAAAGATGGGTGAGAGCTTAT 7496
QY 4974 cctactaagaagagacttcaacggtgtgtgtaagaataatctgaggaatcttccaa 5033
|||||

Db 7497 CCFATAGAAAAGAGACTTCAACCGGTAGCTAAAAAAATCTGAGAAATTTTTCGA 7556
QY 5034 agatttgaatacctaaggaataggttgaagcaatgttccagcttctgttgcagagta 5093
|||||
Db 7557 AGATTGGAATACCTAAGGTAATAGGGTACAAATGGTCCAGCTTTTGTGCCAGGTA 7616
QY 5094 agtcagagactggccaagatlatctgggatttgaatctgaactcattgtcatcagaccc 5153
|||||
Db 7617 AGTCAGAGACTGGCCAAAGTATTGGGATTTGTAATGGAACCTGATTTGATACAGACCC 7676
QY 5154 caaagctcaggaacaggttagagagatgaaatgaacatlaaagaaccccttaacaattg 5213
|||||
Db 7677 CAAGCTCAGACAGGTAGAGAGATGAATGAACCATTAAGAGAACCTTAACTAAATG 7736
QY 5214 accacagaaactggcatatattatgagatgtgtccctccttgccttctttaggtg 5273
|||||
Db 7737 ACCGCGAGACTGGCGTTAATGATTTGGATGATGCTCTCTCCCTTTGTGCTTTTAAAGGTT 7796
QY 5274 aggaacaccccttgagacagtttggcttgacccctatgaattgtcctacaggggaacccc 5333
|||||
Db 7797 AGGAACACCCCTTGACAGATTTGGGCTGACCCCTATGAATTACTACAGGGGACCCCC 7856
QY 5334 ccgttgcagaatgtccttgcacatagtgctgtagtgccttccagccttctgtc 5393
|||||
Db 7857 CCATTGGTGAATAATTCCTTCTGATACATGCTGATGCTCTCTCTCCAGCCTTTGTTTC 7916
QY 5394 tctaggtcaagcgtctgaggtgtgtgagcagcagcgttgagacagctccggagagcc 5453
|||||
Db 7917 TCTAGGCTCAAGGACTTGAATGATGAGTGAACACAGACCTGAGAGCACTCCGGAGGCC 7976
QY 5454 tactc--aggaagagacttgcaagttccacatcgcttccaaattgagatcagatctat 5510
|||||
Db 7977 TACTCAGGAGAGAGACTTGGAGATCCCAATCGTTTCCAGTGGAGATTAAGTCTAC 8036
QY 5511 gtttagagccacggttgcaagaaac 5534
|||||
Db 8037 GTTAGAGGCCACCGTGCAGAGAAC 8060
|||||

RESULT 4
US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TIME OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: Specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
; US-09-111-085-3

Query Match 41.1%; Score 3376.6; DB 3; Length 3482;
Best Local Similarity 99.0%; Pred No. 0;
Matches 3441; Conservative 0; Mismatches 29; Indels 7; Gaps 4;
QY 4718 cctgcagcagttgttgcagaacatccctatcatgttctgaggtcaccagagtggtcga 4777
|||||
Db 6 cctgcagcagttgttgcagaacatccctatcatgttctgaggtcaccagagtggtcga 65
QY 4778 cctgcagcagttgttgcagcagcgtggttaatgtcattcccttccagaaatcttccaa 4837
|||||

Dh 66 ctcgtgtgtcaacatgtgtgtccctgtccagctgtgttaagttaacttccacagataacc 125
Qy 4838 tccaagaagaagactaaagggaaagccaccagcgctcaactcgtggaagtgagcttcacga 4897
Db 126 tccaggaagaagactaaagggaaagccaccagcgctcaactcgtggaagtgagcttcacga 185
Qy 4898 ggtaaagccggctaaataacggaataatctatgtgttttttgaacacctttcag 4957
Db 186 ggtaaagccggctaaataacggaataatctatgtgttttttgaacacctttcag 245
Qy 4958 atggttaagagcttatcctactaagaagaagactccaacgtgtgtgtcgaagaataact 5017
Db 246 atggttaagagcttatcctactaagaagaagactccaacgtgtgtgtcgaagaataact 305
Qy 5018 ggaagaaattttccaagatttgtaatacctaagtaataaggtgacagaatgttcagc 5077
Db 306 ggaagaaattttccaagatttgtaatacctaagtaataaggtgacagaatgttcagc 365
Qy 5078 ttctgtgccaggttaagtcagggagctgtgccaagatatltgggattgtatgtgaacatgca 5137
Db 366 ttctgtgccaggttaagtcagggagctgtgccaagatatltgggattgtatgtgaacatgca 425
Qy 5138 ttgtgcatacagaccccccaagctcagacagagttagagagaatgaataaccataaaga 5197
Db 426 ttgtgcatacagaccccccaagctcagacagagttagagagaatgaataaccataaaga 485
Qy 5198 gaccttaaccaattgacacagagactgtgcatatgaattgattgtgtctccctccctc 5257
Db 486 gaccttaaccaattgacacagagactgtgcatatgaattgattgtgtctccctccctc 545
Qy 5258 tgtgctttttaggggtgaagaaacccctgtgacaglttgggctgtgacccccctataatgtct 5317
Db 546 tgtgctttttaggggtgaagaaacccctgtgacaglttgggctgtgacccccctataatgtct 605
Qy 5318 ctaagggggagaccccccccgctgttgagaaatgtgctttgtcacatatgtctatgtgtctc 5377
Db 606 ctaagggggagaccccccccgctgttgagaaatgtgctttgtcacatatgtctatgtgtctc 665
Qy 5378 ttccaagccttgttctctctagagcagcgctcgaatgtgtgagagagagagctgtga 5437
Db 666 ttccaagccttgttctctctagagcagcgctcgaatgtgtgagagagagagctgtga 725
Qy 5438 gcaagctccgggaagcctactcaag -- aggaagctgtgcaagttccacatcgcttccaagt 5494
Db 726 gcaagctccgggaagcctactcaagagagagactgtgcaagttccacatcgcttccaagt 785
Qy 5495 tggagatttcagctatgttgaacgcacccgtgtgacaggaacactcgaagctcgtgtggaagg 5554
Db 786 tggagatttcagctatgttgaacgcacccgtgtgacaggaacactcgaagctcgtgtggaagg 845
Qy 5555 acctatctcgtacttcttgacacacacacacagcgctgtgaagttcgaaggaatcccccactg 5614
Db 846 acctatctcgtacttcttgacacacacacacagcgctgtgaagttcgaaggaatcccccactg 905
Qy 5615 gatccatgtcatccacagtttaagccggcgacacactcccgactcgggtgtgagagccgaaaa 5674
Db 906 gatccatgtcatccacagtttaagccggcgacacactcccgactcgggtgtgagagccgaaaa 965
Qy 5675 gactgagaatcccttaagcttcgctccacatcgctgtgttcttactacttaacataactc 5734
Db 966 gactgagaatcccttaagcttcgctccacatcgctgtgttcttactacttaacataactc 1025
Qy 5735 cccagagccagtagtaaacgcttataagacagctcgaaccccccatagaaccttaccctta 5794
Db 1026 cccagagccagtagtaaacgcttataagacagctcgaaccccccatagaaccttaccctta 1085
Qy 5795 cctgtgcgtatattgacacctgataaggtgttcaactgttaaatagcaactcgaagtggtgtc 5854
Db 1086 cctgtgcgtatattgacacctgataaggtgttcaactgttaaatagcaactcgaagtggtgtc 1145
Qy 5855 ctaagagcaacctgtgtgctggaactgtatcttgcctcgaattgataaaccccgctgtta 5914
Db 1146 ctaagagcaacctgtgtgctggaactgtatcttgcctcgaattgataaaccccgctgtta 1205

Qy 5915 aaagcacacctcccaactagtcctgtagttagtggtttctatctgtgtccagcagacagaga 5974
Db 1206 aaagcacacctcccaactagtcctgtagttagtggtttctatctgtgtccagcagacagaga 1265
Qy 5975 aaagaaatactgtgtgggtgttcgtgggaacctctctgttaagaatgagctgtcaact 6034
Db 1266 aaagaaatactgtgtgggtgttcgtgggaacctctctgttaagaatgagctgtcaact 1325
Qy 6035 ccaacgattgagacatggaatgtgcccgaatctctctccagacccgggttaaaattctccttg 6094
Db 1326 ccaacgattgagacatggaatgtgcccgaatctctctccagacccgggttaaaattctccttg 1385
Qy 6095 tcaattccggcccggtgggaatgacaaatgtatgaactatataaagaatgaagctgtcc 6154
Db 1386 tcaattccggcccggtgggaatgacaaatgtatgaactatataaagaatgaagctgtcc 1445
Qy 6155 catcagacttagatlatctaaagataagtttcactg -- aaaggaataacaggaataatctc 6212
Db 1446 catcagacttagatlatctaaagataagtttcaactgtgaataaaggaataacaggaataatctc 1505
Qy 6213 aaaagtgtgataaattgttatgagctgtgggaatgttttat - tatgtgcggggagagcag 6271
Db 1506 aaaagtgtgataaattgttatgagctgtgggaatgttttatataataatgtgcggggagagcag 1565
Qy 6272 ggtccactttaaccattgccttagagtagagacggggagacacccctgtgtgcaatgtg 6331
Db 1566 ggtccactttaaccattgccttagagtagagacggggagacacccctgtgtgcaatgtg 1625
Qy 6332 gaccgataaagactgtgtgtgaacagggcccccgccctgtgagccacgcataactgtc 6391
Db 1626 gaccgataaagactgtgtgtgaacagggcccccgccctgtgagccacgcataactgtc 1685
Qy 6392 cgtgtccccaattaaactcgtgtgcgtcgtgacataacacagccgcttagcaacagtaacca 6451
Db 1686 cgtgtccccaattaaactcgtgtgcgtcgtgacataacacagccgcttagcaacagtaacca 1745
Qy 6452 ctgtgttatttccctacacacacgctcagaaaccccccaagtggttccgtttaagaacggac 6511
Db 1746 ctgtgttatttccctacacacacgctcagaaaccccccaagtggttccgtttaagaacggac 1805
Qy 6512 agagacttctcagcttctcaccacagggagcttcccaagcctacactccacagccctgag 6571
Db 1806 agagacttctcagcttctcaccacagggagcttcccaagcctacactccacagccctgag 1865
Qy 6572 ccaacttctctgtgtgtgtgttctatccctcaggggctcccttaattatgaaggagtgcta 6631
Db 1866 ccaacttctctgtgtgtgtgttctatccctcaggggctcccttaattatgaaggagtgcta 1925
Qy 6632 aagaagaagaattcaatgtgacccaagagcatagaataatgtacatgtgggttccgga 6691
Db 1926 aagaagaagaattcaatgtgacccaagagcatagaataatgtacatgtgggttccgga 1985
Qy 6692 ataagcttaacctcacttgaagtttccgggaagggagacatgca taggaagaagctccocat 6751
Db 1986 ataagcttaacctcacttgaagtttccgggaagggagacatgca taggaagaagctccocat 2045
Qy 6752 ccacacaaaccttgtctatagctatgtgtttatgaagagcctccagaataatcagttat 6811
Db 2046 ccacacaaaccttgtctatagctatgtgtttatgaagagcctccagaataatcagttat 2105
Qy 6812 tagtacctgttataacaggtgtgtgacatgacatltggtttaaacccttgtttcca 6871
Db 2106 tagtacctgttataacaggtgtgtgacatgacatltggtttaaacccttgtttcca 2165
Qy 6872 cctcagcttccaacaaatcccaagaatltctgtgtcatgtgtccaaatcgttcccccagtg 6931
Db 2166 cctcagcttccaacaaatcccaagaatltctgtgtcatgtgtccaaatcgttcccccagtg 2225
Qy 6932 actacatctctgaagaagtggtccttgtatgaatagactatcgtgtataacgacaaana 6991
Db 2226 actacatctctgaagaagtggtccttgtatgaatagactatcgtgtataacgacaaana 2285

OY	2845	gaagagtaacgagactttagagagaggttcaataaaagggtgcaagatctaaccccaacgagttcc	2904
OY	2845		
Db	2970	gcccggtacaaagacctccgggaagttatataaaagggtctcggaacattccacctaacagttcc	3023
OY	2905	cgaacctataacctcttagagccccctccgcgcttbaacggagacttgcatacagtatatg	2964
Db	3030	cgaaaccttacaacttacttaagctctctccaccacgagaaaccttgatataggtctctgg	3083
OY	2965	acttaaaagatgcctctctctgcctgagatataccccactagccaacctatttgcct	3024
Db	3090	attttaaagagctctcttctttggctcgctgctgcacccccaaagatcaacttctgtttgcct	3149
OY	3025	tcgaattagagagatccagagtaacgaggaacacggcgaactcaactctgcgacgtgcgcctcc	3084
Db	3150	ttgaatgagagagaccacagagggcgagactggtcaatttaaccttgagactatacactacac	3203
OY	3085	aagagttcaaaactcccgcgacctctttgaagaagccctacaagaagagacgtgcgcaact	3144
Db	3210	aggggtttcaaaaactcccccttaacctgttttagacaaacctccatgcgggatttagaacctt	3265
OY	3145	tcagatatacaaacacctcaagltgacacctccctccagtaacgttgatgacctgcctgcg	3204
Db	3270	ttccgggctcaaaaacccccagotcttaacctactgacagtagacgtatgttttaatgcag	3322
OY	3205	gaagccaacaacagagactgcttgaagaagtaacgaagcactactgtgaaattgtctgac	3264
Db	3330	ccgcctcaaaaagacctatgttcaacaggggactagagactccttcacagaaactggggatt	3389
OY	3265	taggctacaagagctctgcataagaagccccagatttgcagagagagagtaacatctgg	3324
Db	3380	tggggtatccagtttcgggtcaaaaaggcacaaatttgcataacgtaggtatcttctgg	3449
OY	3325	gttcaagtttgcgagcgaggacagatgctgtgcacgagagacagaaagaactgtagctc	3384
Db	3450	ggttatcttcttcgagagaggtataaaagattgctcactgaacccccgaaaaaagactgtc	3509
OY	3385	agataccggccccacaacacagccaagaatgagagatttttggggagacgtgtatltt	3444
Db	3510	tgaattccaccccccacacacacacagcgaggtacgtatgatttttggggagactgtcgtctt	3569
OY	3445	gcagactgtgattcccgaggttttgcagacttaacccccactctacccgctcaaccaag	3504
Db	3570	gtagactcttgatctcaggtcttggccgacttggacacctacacacctctatgtcccttgactacg	3629
OY	3505	aaaaagggggaattctccctggcctccctgcagacagaagaagcatltgatctatacaaaag	3564
Db	3630	aaggaatttccttggatggaagaagaagaaacacaaagaccttttgagcgtatcaaatgt	3689
OY	3555	ccctgtcgtagcgcacactgctctgcgcccctccctgaacglaaataaaccttacccttatg	3624
Db	3690	ctcttaatgactgccccctgcctctgacatttaccacacttggactaaagctcttcgtctaat	3749
OY	3625	tgatgtgacgtgaagaagagtagcccgagagattttaacccaaccttagacacatagaga	3684
Db	3730	tggacgaaagagccgggatagccagagagggctgtgacacaaacacttagaacaccttgaaaga	3809
OY	3685	gacctgttgcctaccttctcaagaagaacttgatcctctttagcagagtggtggccgatatctc	3744
Db	3810	gaccggtagacctatctgtcaaaaaaactggaatccgctgtcgtatgtgagtgccacatatgtc	3869
OY	3745	tgaagctatacgcagcgtgtggtgcatactgtgtcaagaacgtgcgaataatgactttggag	3804
Db	3870	tgaagagctatttgcagcagtagtgcctctctgtatcaaaantcctatpaaactgacaaatgggac	3929
OY	3805	agaataaactcgtlaaataagcccccaactgcatlttgagaagaactgttgggaagccccagacc	3864
Db	3930	agcaggtactgtctgtggccccctctatgacccttgagagatatttgcggcagccacctgtata	3999
OY	3865	gatgtatgacaaagccccgatgacacacataaagccctgtcttctcaagaagaaaggtca	3924
Db	3990	gatgatgactaaatggccccgatgacacacactatgacagctgtcgtcttaatatgacgggata	4049
OY	3925	cttgcgtcccaacagccgctctcaacctgtgcaacttctgccttgaagagtgtatgac	3984

Db	4050	CGTTGGCCCCCCTGCGCATTTCTCAACCCAGCCACCTCCTCCTCTAAAGAAATGATTCGG	4106
Qy	3985	cagtgacatacgaatctgcacatcaaatatgtatgtaggagagactgggtgcgcaagaactta	4044
Db	4110	TCCCGATACATTCGATGTGCGAGACATCCTGGCTGAGAAATTTGGGACCAAGAAAGCCTGA	4165
Qy	4045	cagacatacgcgtactgtaggaagtgctaaacctggtctactgaacgaagcagactgtg	4104
Db	4170	CTGACCAACCCCTGGCCCTGG--ACGCTCTAGCTGGTACACGGAATGGCAGACTTTTCTGA	4226
Qy	4105	tggaaagttaagaagatggtgtggggcgagtggtggagcggaaccgcagactctggcca	4166
Db	4227	TTGAGGGAAACGGAAGGCGCTGAGACTGCGGTGTGATGTGAAAAAGTAAATTTGGCA	4286
Qy	4165	gcacgctccggaaggaacttcacgcgaagaagctgactgacatgagccctcagcaagct	4224
Db	4287	GTGCGCTTGCCCTGGAAGGAACCTTGGGCACAGAGGCTGACTCATAGGGCTTACACAGGCC	4348
Qy	4225	tgcggtcggccgaagggaatccataaacttatacggacagcagtaigtcccttgga	4284
Db	4347	TCCGAGAGCGCGAAGGTAAGATCATTTAAATTTACCTGACACGCCCTATGCTTTTGCTA	4400
Qy	4285	ctgcacacgtaacacggggccatctatataaagaagggtgtgcttacctagcaaggag	4344
Db	4407	CTGCACATATCTCAATGGGGCCATCTACAGCGAGCGAGGTTATTTGACTTTCAGCGGTAAAG	4466
Qy	4345	aaataagaacaagaaggaatcttaagcctacttaagaaccttactcttgcgaagaagc	4404
Db	4467	ACATCAAAACCAAGGAAGAAATTTCTGGCCCTGTTAGAACCATACATGCGCCCAAGAAAG	4526
Qy	4405	tagctatatacactgtctctggaacatcagaagaaccaaagatcatalctagaaggaa	4464
Db	4527	TAGCCATCATCTCCTGTCGCCGTACACAGAAAGAGACTTAGTGCCCAAGGGCAACC	4586
Qy	4465	agatggtctgaccgggttgccaagcgagccagctgttaacctctgcctataatag	4524
Db	4587	GGATGGCAGACTCGTAGCAAAAGAGTTGCTCAGGGGCCATGATCTTAACTGAAAAAG	4646
Qy	4525	aaagcccaagaagccccaagaaccgaagacagatcacaccttagaagactgycsaagataa	4584
Db	4647	GAAATCC-----GTCCAAAGCCCTGA-----GGATGAAACT	4679
Qy	4585	aaaagatagaccagttctctgagactccggaaggagacctgctataacctataatggaa	4644
Db	4680	ATGATATTAAGAGACTATTTTGGACTAGTATCCCTCCCATACTTTTCGAAGGAAAA	4739
Qy	4645	aaatctgcgcccacaagaagggttagaataatgtccacaagatcatgtctcaaccac	4704
Db	4740	TAGATTTGACACCGCAGAGAGGAATAAAATTTGTGAAGAGACTACACCGATTACCCAC	4799
Qy	4705	taggaactaaacacctgcacagatgtgttagaacaatccacctatcatgtcttgaggctac	4764
Db	4800	TGGGACTGAAAAAATATATAGACTGATTTAAAAAGTCCGGTATCAAGTCCCTAACTTGA	4855
Qy	4765	caggagtggtgactcgtgtgtlcaaacatgtgtgccttcgcagctggttaatgtlaac	4824
Db	4860	AGTCAGTGGCTCAAAAGATTATTAACCTCTGCAAAACGTTGCAATTCACATAATGCACTA	4919
Qy	4825	cttcagaatacctccagaagaagagactaaagggaagccaaccagcgctcacttgggaag	4884
Db	4920	AAACCTTACAAAGAACCTGGGAAGAAGACAAACGGGGAACCGTCTCGAGATGTTGGGAG	4979
Qy	4885	tggacttcaactgaagttaaagccggctaaataaaggaacaatatcatgtgttttgtag	4944
Db	4980	TGGATTTTACTAGAACTTAAACCCCGGAATGATGTGTAATTAAGTATCTGTAGTATTGTAG	5039
Qy	4945	acacctttcagaatggtgttagaggtatctactactaagaagaagacttcaacctgtgtg	5004
Db	5040	ACACCTTTTACGATGGGTGTAAGCGTTCGCCACTAAACCTGAGACTGCCAGATTGGG	5099
Qy	5005	ctaaagaaatacttggaggaaatcttccaaagatttggaaacctaaagtaagtaaggctag	5064

Dh 5100 CCAAGAGATTGTAAGAAATCCCTGCCAAGATGAGTACTTAAGGTAATCGGTCGG 5159
Qy 5065 acaatgtccagcttcctgtcccaaglaagtcaggagactcgcacagatattgggagatg 5124
Dh 5160 ACAATGAGACCAAGCTTGTGGCCAGGTAAGTACAGGATTTGGCCACTCACTTGGCATTCG 5219
Qy 5125 attgnaactgcatltgtacatacagaccaccaagctcaggaacggttagagagatgata 5184
Dh 5220 ATTGAATTAACATTTGCTTACCGCCCTCAAAAGCTCAGACAGGTAGAGAGATGATA 5279
Qy 5185 gaaccatlaagagagacccatlaacatattgacacagagacgtgcatlaagatgagag 5244
Dh 5280 GAACATTAAGAGACACTTGACTTAATTTGGCCATGAGAGACGGCGGGAAGAGACTGGGTGG 5339
Qy 5245 ctccctgcgcttctgtctttagggtagagaaaccccttgagacaglttgagctgcaccc 5304
Dh 5340 CTCCTCCCTGCTTGGCCTCTTCCGAGCCGGAACACCCCGGACGTTTGGGCTCACTC 5399
Qy 5305 cctatgaattgcttaagggagagacccccccgttgacagaaatg-----ccttgac 5358
Dh 5400 CTTTGAAGTCTGTATGGGGGACCTCCCTTAATTAAGATGTTGGAACATTTGGTTTC 5459
Qy 5359 atagtcgtatgctgtcttccagccttctctcctaggtctcaagcgctcagatgag 5418
Dh 5460 CGGATTCAGGCTCTGTCTTACCTCCTCTTGTATTCATTTAAGGCCCTGAAGTAA 5519
Qy 5419 tgaagcagcagcgtagaagcagctcggagagcctactcagagagagacttgcaatg 5478
Dh 5520 TTAGAGCCCAATTTGGGACACAGCTAAGAGAGGCTTACACCCACGAGACCCGACATG 5579
Qy 5479 cacatcgcttcccaaglttagagattcagctctatgtagacgcacacgttcaagaaactc 5538
Dh 5580 CCCAGAGATTCCAGGTCGGGAGACCAAGCTTGTGTAGAGACGATCGAGATCGGTCGTG 5639
Qy 5539 agactcggtagaagagacattctcgtactttgacacacacagcgctgtgaaatcg 5598
Dh 5640 AACCAAGGTGAAGGAGACCTTATTAAGTGTACTAACAACTCTTACGCGATGAAGTGTG 5699
Qy 5599 aaggaatccccacgtgatacgtatccacgttgaacggcgagc----- 5644
Dh 5700 ACGGATTTGCTCTGAGATCCACGCTTACACAGGTCAAGAGGCCCCCAAGTCAAGTGAAG 5759
Qy 5645 -acctccagactcggggtgagagccgaaagactgagaatcccccttaagcttgcctcc 5703
Dh 5760 AAACCCAGCAAGAACACTGGGCTGTGGAAGCCACTGATACCCCTTAAAGCTTGCCTTCG 5819
Qy 5704 atcgccgtgttctacttacttaacat----- 5729
Dh 5820 GTGCGAGAGAGCCCTCCATCCACCTGAGACTAGGGAACACACCCCTCATGCCCCAGTT 5879
Qy 5730 -----aaactcccgagcgagtagtaaacgccttataagacg-----tc 5768
Dh 5880 CAACACTCTTGGGAAGTGTCTTAATGAAGAGAGACGTTGTATGGGTAGCCACTGCAGTC 5939
Qy 5769 gaaccaccaatagact-----ttatcccttaacctgctgattat 5807
Dh 5940 CATCCCTTTGGACTTGTGGGCTGATCTCACACCTGACATCTGTAATTAATACACAGGA 5999
Qy 5808 tgaacctgatacgggtgtcactgtataatagcaactcgaaggtgtgtcctcctagagacgt 5867
Dh 6000 TCTCCCAATTTGGGACCTTCCGATCACTACTGACTTGAATTAACCCACCTCTGAACAAAAG 6059
Qy 5868 gtgagcctgaagcattcttgcctcgtatgattaaacccgctgttaaaagacacccctc 5927
Dh 6060 TGTGTCCCAACAGGGGTAGGAAGCACTACTGATGTTCCGGGCAAGTTTATGAGAGTAT 6119
Qy 5928 caactagtcgtagtataaggttctactgtcgtccagacagagaaa-----gagaaa 5982
Dh 6120 CTTAGGGGCTGCACAGTTTATGTTTGCCTGTGTCAGGGTCAGAAAGAAAGCTGCAACAA 6179
Qy 5983 tactgtggggttctcgtgggaactctctgttagagagtagagctgtcctcccaacagat 6042
Dh 6180 GAATGTAGAGGGGCACTCACTTCTTGGGGTAATGGACATGTGAACACAGGGGGA 6239

Qy 6043 ggaagctgaaa-----atggcattctctccagagccggttaaaattctccttgtcaa 6098
Dh 6240 GCTTACTGGAAGGCTTCCCTGATTTGGAGACCTGATCAGGTAAACGTGGTATGTTAT 6299
Qy 6099 ttccggcccggtgcaagtaacaaatgataaactatataaagataagagctctcccatc 6158
Dh 6300 GATTAACCAACCAAGAGAAAGAAACCCTCAATTAATTAATTAATTAATTAATTAATTA 6359
Qy 6159 agacttagattatcctaagataagtttcaactgaagaaagaaataatc--aaaa 6216
Dh 6360 AAAAATTAACAGCCCCCAGAGACCATGCAAGTAAATTAATTAATTAATTAATTAATTA 6419
Qy 6217 gtgataaattgtagtagctgtgggaaatgttttattatataatgtcggggagagagctc 6276
Dh 6420 TTACTGAGAAAGAGAAACAGCTCCCTGAGTGTGGCTTAAAGAAATAGTGGGGTTGG 6479
Qy 6277 acttaacattgccttagatagagacgggagagacccctgtgcaatgagacc 6336
Dh 6480 CGAGTATTAATTAATTAATTAAGAGACCTGGGTTTATCTTAACTTAAGACTGAAGTAA 6539
Qy 6337 gataaagttatgctgtgaacagggcccgccctgagacacagcataactgtccggtg 6396
Dh 6540 GACCTGGCACTTAATCATTAGAGACCAACAGGTCCTTAAGAGAGAGAGAGAGAGAGT 6599
Qy 6397 ccccaatlaacctgcgtcgtgagacataacacagcgctagacacagtlaccatgga 6456
Dh 6600 GCACCGGCTCCCGGAGAGTCCAGCGGTCCAGCTCCACCACTTACAGGCGCTTACACA 6659
Qy 6457 ttagtctcctcaacacagccttagaaactccccaggggttctcgttlaagacagagaga 6516
Dh 6660 GTAGAGACCTTATTAAGAGACACCTTAAGCTCCCAACACTCTTAATTAAGAGAGAG 6719
Qy 6517 ctcttaagctcactcagagagcttccaaagcactcaactcaacccctgtagcact 6576
Dh 6720 CTGCTGAGTGTAGTTCAGGAGGCTTTTGTATTAATTAAGAGTAACTTAATTAATTA 6779
Qy 6577 tctctgtgtgctgttctatcctcagagcctccttattatagagagtagtgaagaa 6636
Dh 6780 CAATCATGTTGTTATCTATGCTTACACCCCTTATTAAGAGAGAGAGAGAGTGTGAGCT 6639
Qy 6637 agaaattcaatgtgacacaaagagacatagaatcaatgtaacatgtaggggtcccaataag 6696
Dh 6640 AGAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6896
Qy 6697 cttaacctcaatgaagttccgggaagggacatgacatagaaagagctcccccatccac 6756
Dh 6897 TTGACTGTGACAGAGTTCAGGAAGAGGCTTGTAGGTCAAGTGGTCCGAGATGA 6956
Qy 6757 caacacctgtgctatagtagctgtgttataagagagccctcagaaataatgattatgta 6816
Dh 6957 TGGACCTCTTACACAGACTCAAAATTAATTCGACCTTAACAAAGGTGTGATCTGATG 7016
Qy 6817 ccttgattaaacaggtgtgtggaatgcaatagctgtgttlaacccccctgtgttccacatca 6876
Dh 7017 CTTCCCATAGACACAGTATGGGCTTGCATTAAGAGTCTCACTCTTGAATATCATGTC 7076
Qy 6877 gtcttcaaccaatccaaagattctgtgtcattgltccaaatcgtcccccgagtgatctac 6936
Dh 7077 GTTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7136
Qy 6937 catcttgaggaaggtgtccttgatagataatgactatgataaccccaaaaagagaa 6996
Dh 7137 CATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7193
Qy 6997 cccgtatcccttaacctagcgtgtaatgtcgtgalt--aggagagcgtgtgtgagta 7053
Dh 7194 CCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7253
Qy 7054 acaaggacagctgcccgtatcaacagacacagcagcttagagaaagactgtgtgagta 7113
Dh 7254 ACTGGAACCGCTGCTTATATCCAGACCCCGGATTAATTAATTAATTAATTAATTAATTA 7301


```

QY 7114 catgagccatgacagaagatctccagccttaagaagatctgttagcaacctagaagag 7173
Db 7302 GGTACAGCTATGATGACTGATCTCAGACCTTAGAACACTATTAACCAACTAGAGAA 7361
QY 7114 tcccgactctctgtcgaagtggtctcaagaacgggggagatagatctcgctgtt 7233
Db 7362 TCTTAATCTCTGTCTGAGAGTACTGCAAAATAGAGAGAGATTAAGATTGTTATTT 7421
QY 7234 ctaagaaggttggttattgtcagccttaagaagaatgtgtctctctatagatcgc 7293
Db 7422 CTTAAAGAGAGAGACTTTGTGCTGCTCAAGAGAGAACTGTTTATATGTCGACCC 7481
QY 7294 tcaagagccatcagaagactcatalgaagaacattagaagaagtgagtgagag 7353
Db 7482 TCGGAGAGATCAGAAAGTTATGCGCAACTTAGAGAAAGCCCTATATAGCTCAAGA 7541
QY 7354 gaaagagaggtcagcaggggtgttgaagaatgtgtcaagagctctctctgagac 7413
Db 7542 GAGAGAGAGAGCAACAAAGATGTTGAAGAGCTGTTATATAGTCCCTGCTGACCC 7601
QY 7414 accctgtctctgtcgtcgaagggccctagtagtctctctctctacttaagttgag 7473
Db 7602 ACTCTCTTCCACTATAGGCGGACCCCTGATATATCTGCTTTGCTTACTTTTGGC 7661
QY 7474 cctgtctaataataggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 7533
Db 7662 CCGTGCATTTCTTATATAGTACTGCTTTATTTAGAGAAATTAATGCACTACAGGTT 7721
QY 7534 atgtacttaggcaacagtaacagagcctctg 7566
Db 7722 ATGTTATTAACACAAATATCATGCTTCCAG 7754

```

RESULT 6

```

US-08-716-351A-1
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:

```

```

; APPLICANT:
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/716.351A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-128-1PC
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; MOLECULE TYPE: DNA (genomic)
; HYPOHETICAL: NO
; FEATURE:

```

```

; NAME/KEY: misc.feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name= "Galy SEATO Genome"

```

US-08-716-351A-1

```

Query Match      27.3%; Score 2244.4; DB 3; Length 8535;
Best Local Similarity 59.1%; Pred. No. 0;
Matches 4366; Conservative 0; Mismatches 2811; Indels 215; Gaps 22;

```

```

QY 332 agcagagctctgaagagatcacaagctcgaaccttgaggagagcccgaggttgagga 391
Db 756 AACAGAGGTGTTCGGGGCTCACCGCCCTGTAAATCCAGGGAGAGAGCTCCAGATCAGGGGA 815
QY 392 gaggcagagagccttggtgtgtcctcactgtgtgtaagagagccaggtctgttga 451
Db 816 GGACCAAGAGAGCGCTGTGAGACCCCTCG-----GTAAAGGCTGTGTATACCG 864
QY 452 gcaagaagctccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 511
Db 865 ATTTGATCGCCCTGTTGTAAGACGCGCTGTGAATCTGATTCMTCTCGGTCCGCTCG 924
QY 512 ggttcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 571
Db 925 CCGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 984
QY 572 tgcataagtttaataatgagagagagagagagagagagagagagagagagagagagag 631
Db 985 CTCCTCAATATCATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1044
QY 632 ttgactgaagtttagatcagaaggtcacaatttgtaagttcaggttaagaagagacctg 691
Db 1045 CTGAGAGATGTTGAGAACAAAGGGCTCACATCTATCCGTGGAATCAAAAAGGGAATG 1104
QY 692 gcagactctctgtcctcgaatgagagagagagagagagagagagagagagagagagagag 751
Db 1105 GCAGACTTTCTGTCTCTCGAGTGGCCCAATTCGCGGGGCGGCGGAGAGAGAGAGAGAG 1164
QY 752 cttaattcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 811
Db 1165 TTTTAATCTCTCTGATTTTGTGAGTTAAAGATGTCTTTGAG---GAGAACGGGG 1221
QY 812 tcatctgaagagagagagagagagagagagagagagagagagagagagagagagagagag 871
Db 1222 ACATCCGAGCAAGTTCCATATATGCTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
QY 872 atgggttaaacatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 931
Db 1282 ATGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1341
QY 932 agagagagagagagagagagagagagagagagagagagagagagagagagagagagag 991
Db 1342 GAGGCCATCAGCTCTCCCGACCCCGACCCCGACCCCGACCCCGACCCCGACCCCGACCC 1401
QY 992 tcgagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1049
Db 1402 CTCGTACACCAAGCCCGCCCGCTTATCCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1461
QY 1050 acaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1108
Db 1462 GATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1165
QY 1109 tgcagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1165
Db 1522 GACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1581
QY 1166 ataccgtgcagacatcagagagagagagagagagagagagagagagagagagagagagag 1225
Db 1582 TTTGCCCTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1641
QY 1226 ttgaccttctctgtcagatctataatgtgaagaacacacacacacacacacacacacac 1285
Db 1642 TTGGCTTTTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1701
QY 1286 ggaatcccaagagagagagagagagagagagagagagagagagagagagagagagagag 1345

```

Dh 1702 AAACCCAGCAGCTCTACGGGGCTCTTGAAGTCTTATGTTCTCCCATCATGCCACTG 1761
Qy 1346 ggaatgattgcacagctgctgcagacaccltcacacccagagagagaaatctc 1405
Dh 1762 GGAGATGTGACACAGCTCCCTACAGATTCTTTTCCACCTGAGAGACGGGAAGAAATCT 1821
Qy 1406 gttgaggtctagaaaaatgtctcctgggagcgacggcgagccacgagttgcaaaatga 1465
Dh 1822 CCTGGAGGCCCGCAAAATGTCTTGGGGACAAATGGGGCCCTACACAGCTCGAGACCT 1881
Qy 1466 gattgacatggattcttcctgactgcgcccgggttggagctacacacggctgaagtag 1525
Dh 1882 CATTAATGAGGCTTCCCTCAATGCACCTCACTGGGATTTACAAACACCCGAGTAG 1941
Qy 1526 ggaagctctgaaatctacacgagctcgtgtggtggtgtccggggcgccctaaagag 1585
Dh 1942 GGAGGCTTCTGTGCTTACCGCCGACTCTAGTGGCAGGCTTCAAGGGGCGAGCTCGGCG 2001
Qy 1586 gcccaactaatttgctaaagtaagagagtgatgcaaggacgagacgacccctcctgct 1645
Dh 2002 TCTACCAATTTGGCTAAGGTAAAGAGAGGTCTTCAGAGGACCGGACAGACCCCTTGCT 2061
Qy 1646 attcttgagaggtctcatagagacctcagcggttcaaccctttgactcactagata 1705
Dh 2062 TTTCTTGAACGCTGTGATGAGGCTTATAGAGATACACTCCGTTTATCCCTTCTTGA 2121
Qy 1706 ggcacagaagacccctcagtgccctgcccctcatttggtgagctgagctctgatacagaa 1765
Dh 2122 GGGACACAGGCTTGGCTGCTGAGGCTCCAGGACTTCTTCAAGATTTAGTAAAGAGGC 2181
Qy 1766 gaaactcagagactggaaggtgtacagagagctgagttacgtatcagtagtgagagagc 1825
Dh 2182 AAATTTACAGGCTTGAAGGCTCCAGGACTTCTTCAAGATTTAGTAAAGAGGC 2241
Qy 1826 agagaaggtgtattacag 1885
Dh 2242 AGAAAGGCTGTACATAGAGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2301
Qy 1886 gagaagaagaag 1945
Dh 2302 GGGAG 2361
Qy 1946 ggcgcagtggt--tgaag 2003
Dh 2362 GGGCGAGTAGTAGTAG 2421
Qy 2004 aggtcagagccctagacagctcaggaagacgtggtgcaatagagacccactcgaagagcag 2063
Dh 2422 GGCAGAGAGAGACCTTGGGATGGA-----AGACCTCCACTAGCAAGAGACAG 2470
Qy 2064 tgtgcatgttgaagaagaagaag 2120
Dh 2471 TGCCATACCTGTAAAG 2530
Qy 2121 aagaagacgaaggtcctagctctagagaagaatagaagattagagagagagagagagagag 2180
Dh 2531 AGAGAGCCCAAGGTTTACGCCCTAG-----ATTAACTAGGGGAGTCCGGGTTGGGAG 2581
Qy 2181 cccctcccgagcccaaggttaacttgaaggtgagagagagagagagagagagagagagag 2240
Dh 2582 CCCCCTCCCAACCTAGGATACACTGACTGTGAGAGGAGCCGCCCATTTGATGCTCTGGTTC 2641
Qy 2241 gatacag 2300
Dh 2642 GACACCGGAGCTGACATTTAGTATTTAGCCCAACCCATGGGAAAGTAGGGGCTCCAGACGG 2701
Qy 2301 tcttgagtgatggtgacacag 2360
Dh 2702 ACGGTCTGAG 2761
Qy 2361 gacttgag 2420
Dh 2762 AAAATTGGACATAAACAAGTACCCACTCTTCTGTGTATACCGAGTGCCTTGTCTGT 2821

Qy 2421 ctctagtagagacttactagaccaaagatgagagagagagagagagagagagagagagagag 2480
Dh 2822 CTGTTGGGAGGAGACCTCTCAACCAACCTAAAGGCCAGATCCAGTTTCCGTGAGGGC 2881
Qy 2481 ccagaagtgctgtgataaacaacccaactcactggtgtgacccctccatlagatgata 2540
Dh 2882 CCACAGGTAAACATGGGAGAGAACGCTTACTATGTGCTGCTCTTAACCTGGAAGAGAA 2941
Qy 2541 taccgactatctccccaagtaagagctgtatcaagatatacaggtccctggttggagag 2600
Dh 2942 TACCGACTACATGAAGAACGAGTACCTCTCTATGCA--CCCATCTCGGTCCAGCTT 2998
Qy 2601 ttccccaagccctgggacagaag 2660
Dh 2999 TTCCCACTGTATGGGACAGAAAGCGCCGATGAGACTGACCAATCAATCCACCAAG 3058
Qy 2661 gttatcaactggaagccaggtgtacacagatatacagacagtagtacccttggataga 2720
Dh 3059 GTAGTAGAGCTAAGATCAGGTGCTTCCACAGTGGCTTTCGCAATATCCATATGAGCAAA 3118
Qy 2721 gaggctcag 2780
Dh 3119 GAAAGCTGGGAGAGTATGAGACCCACATCCAGAAAGTCTTAGACCTAGGGGCTTGGTG 3178
Qy 2781 cctgtccaatcccttggatatacctccctgctacaggtttaggaagccttgagacaaatgat 2840
Dh 3179 CCTGTGCGGCTCGCCCTGGATACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3238
Qy 2841 talcagaccagtagagagacttgagagagagagagagagagagagagagagagagagagag 2900
Dh 3239 TATCGGCAAGTTCAAGACCTGAGAGAAATTAATTAAGAGGTACAGATTTATCCACCA 3298
Qy 2901 gtcgcgaacccctataactcttgagagcctccctccgctgagacggaactggtatcacagta 2960
Dh 3299 GTCCCAACCCCTTACATCTTCTGAGTTCCTTCCGCTTAGCTATACCTGTGATCAGTC 3358
Qy 2961 ttgagactaaagatgctctctctcctcgtgaattacaccccaactgagacaaactttt 3020
Dh 3359 TTGATCTCAAGAGATCTCTTTTCTGCTCAGGCTACATCCCAACGACGCGCTGTTC 3418
Qy 3021 gcttcgaatlgagagatcagagtagagagagagagagagagagagagagagagagagagag 3080
Dh 3419 GCGTTGAGTGGAGAAAGACCCAGAGAAAGGTAAACACAGTGCAGTCCGAGGGGCTA 3478
Qy 3081 ccccaaggttcaagaactcccgacacactcttgagagagagagagagagagagagagagag 3140
Dh 3479 CCACAAGGCTTCAAGAACTCTCCACTCTTCGACGAGGCGCTCCACCGAGATTTGGCT 3538
Qy 3141 aactcagatcacaacacactcaggtgagacccctcccgagtaggtgagagagagagagag 3200
Dh 3539 CCTTTAGGGCCCTCAACCCCAAGGAGGTGTTACTCCATATGTGAGACACTCTTGGTG 3598
Qy 3201 ggcgagagccacaacag 3260
Dh 3599 GCGGCCCCACATATGAAACACTGCAAAAAAGAGAACAGAACCTCTTACAGSAGTTAAGT 3658
Qy 3261 gacctaagctacagagcctgtctagaagagcccaagatttgagagagagagagagagagag 3320
Dh 3659 AAGTTGGGCTTACCGGTTATCGCTTAAGAGGCCCGAGCTCTGCAAGAGAAATCACTAT 3718
Qy 3321 ttgggttagagatttcgaggggagagagagagagagagagagagagagagagagagagag 3380
Dh 3719 CTGGGCTACTCTCAAG 3778
Qy 3381 gtccagataccgcccacaacag 3440
Dh 3779 ATGAAATATCCCTGTTCTTACAGACCCCAAGAGAGTCCGTAATTTCTAGGACATGCGCGA 3838
Qy 3441 ttctgag 3500
Dh 3839 TTCTGAGGCTCTGATCCCTGGGTTTGTCTCCGCTGACACCTTGTACCCCTTAACA 3898

Db 2046 GTTGCTGCTACTGCTGCTTACGGGACAGACAGATGACAGGAGAGAGCCAAAGAG 2105
QY 2042 gaccccatcgacaagaacagctgtcgatgttaagaanaaagacatgggcaagaa 2101
Db 2106 GCCCACTCCACACGACGACGCTGCTACTGCAAAAGAAAGGACATTTGGCTTAAGA 2165
QY 2102 ctgcccagaagaaggaaacaaagcgaagtccta-----gcttaagaagaa 2152
Db 2166 TTGCCCCAAGAGACAGAGAGACCCCGGGACACAGACCCAGGCTTCCTCTGACCTT 2225
QY 2153 taagaatlagggagagcggtgtcgacccctcccgagcccaagggttaacttgaagt 2212
Db 2226 AGAGATTAAGGAGGTAGGGTCAAGAGACCCCTCGAAGCCAGATTAACCTCAAGT 2285
QY 2213 ggaaggcgacaacagttgaatcctgtgtgataccgagcggaagcatcagtgctgtaca 2272
Db 2286 CGGGGGGACACCCGCTACCTCTGATGATGAGGGGCCCAACACTCCGTGTCGACCA 2345
QY 2273 accatlaggaactaaagaanaaaatcctgggtgatgtgtccacagggcaagga 2332
Db 2346 AAATCTTGACCCCTTAAGTGAAGTCTGCTGGGTCCAGAGGCTACTGGAGGAAAGCG 2405
QY 2333 gttaccatgagactaccggaagacggttgacttggagtgagcggttaaccactgt 2392
Db 2406 GTATCGTGTGACACGAGATCGCGAGTGACACTAGCCACCGTAAAGGTCAACCTCTTT 2465
QY 2393 tctgtatcctcctgagtgagcagtaaccctctagttagagacttactgacaaagt 2452
Db 2466 CCTCATGTACCAATGCTCCCTTCTCTGATGAGAAAGATTTGCTTAACCTAAACCTAA 2525
QY 2453 agtcaaatctcttga---acaaggaagccagaagtgctgtgataacaacat 2509
Db 2526 AGCCCAATTTACTTTGAGGATACAGAGCTCAAGTTGTGGACCAATGGGACAGCCCT 2585
QY 2510 cactgtgtgacccccaatagatgaatgaatgcactatctccccaagtaaac 2569
Db 2586 GCAAGTGTGCTACCTAAACATAGAAAGATGATGCTGCTACATGAGACCTCAAAAGGGCC 2645
QY 2570 tgaatcaagatatcacgtcc---tggtggagcagttccccaagcctggcgagaaccc 2626
Db 2646 AGATGTGCTCTTAAGGTGTCACATGGCTCTGATTTTTTCCAGGCGGGGACAGAAACGG 2705
QY 2627 aggatgtggttggcaagcagttcccccacaggttatcaactgaagccagtgtac 2686
Db 2706 GGGCATGGGGCTGGCGCTTCCCAAGCTCCTCTGATCATACCTTGAAGCAACCTTAC 2765
QY 2687 accagatcagtcagacagtaacccttgaagtagagagctcgagaagaaatttggccga 2746
Db 2766 CCGGTGTGCTCAATAAACAATACCCCATGTGTCAAGAAAGCCAGACTGGGATCAAGCCCA 2825
QY 2747 tgttcaagaatlaalccaagagccatcctcagttccgtccaatcccttggaaatacc 2806
Db 2826 CATACAGAGACTGCTGGATCAGGGAAATTCGTACCTGTCCAGTCCCTGGACACAGCC 2885
QY 2807 ccttcaaccggttaggaagccttggaccaatgatatcagacagtaacagacttgaaga 2866
Db 2886 CCTGCTACCCGTTAAGAAACCGGGGACTAATGATTTAGGCTGTGTCCAGGATCTGAAGA 2945
QY 2867 ggtcaataaaagggtgcagacatacaccgaagcgtccgaagccctataaccttga 2926
Db 2946 AGTCAACAAAGCGGTGGAGACATCCACCCACCGTGCACCAACCTTAACAACCTTTGAG 3005
QY 2927 cgcctcccgctgaagcagctgtgtacaagatattgactttaaagaatgcttctctgt 2986
Db 3006 CGGGCTCCACCGTCCACCAAGTGTACCTGTGACTTAAATAATCTTTTCTG 3065
QY 2987 cctagattacaaccactagccaacacttctgcttgcgaatggagagactcaggttac 3046
Db 3066 CCGTGAATCCACCCACCACTAGTCTCTTGGCTTTGAGGAGAGATCCAGAGAT 3125
QY 3047 gggaagaccgagctacactcgtggaacgagctgcccgaaggttcaagaactcccgac 3106
Db 3126 GGGAAATCTCAGCAATTAACCTTGACCAAGACTCCCGAGGGTTTCAAAAGTCTCCAC 3185
QY 3107 catcttgcgaagccctacacagagccttgcgaacttccagatccacacccatcaggt 3166
Db 3186 CCGTTTGAAGAGCCCTGACAGGAGCCTCGAGACTTCCGATCCAGACCCAGACCT 3245
QY 3167 gaccctcccaatgagtgatgaatcgtcttcggggagagcaaccaacagactgtct 3226
Db 3246 GATTCTGCTCACTATGATGATGACTTACTGCTGGCCGACACTTCTGACCTTGACTGTGA 3305
QY 3227 agaagtlacgaagcactactcgtgaattgttcgaactaggctacagaacctgtacta 3286
Db 3306 ACAAGTAGCGGCCCTGTTACAACCTTAGGGACCTCGGATATCGGGCTCGGCCAA 3365
QY 3287 gaaagcccaagatttgcagagagaggttaacatacttgggttacaagtttcggggcgga 3346
Db 3366 GAAAGGCCAAATTTGCCAAGAACAGGTCAAGTATCGGGGTATCTTAAAGAGGRTA 3425
QY 3347 gcatgtgctgacgagcaacgagaanaactgtagtccagataccgcccacaacagc 3406
Db 3426 GAGATGTGCTGACGAGGCCAGAAAGAGACTGATGGGGCAGGCTACTCGAAGACCC 3485
QY 3407 caaacaagtgaagagttttgggacagctgatttgcagactgtgatacccggtt 3466
Db 3486 TCGCAACTAAGGAGAGTTCCTAGGAGAGGAGGCTCTGCTGCTGATCCCTGGGTT 3545
QY 3467 tggacctagcagggcccaactcaccggttaacaaagaaaagggttccctgggc 3526
Db 3546 TGCAGAAATGGCGCCCTTGTACCTCTACCAAAACGGGAGCTCTGTTGAAGGGG 3605
QY 3527 tccctgagccagaaagcatltgatgtatcaaaagccctgctgagcgacactgtct 3586
Db 3606 CCCAGACGACAAAGGCGCTACCAAGATCAAGACGCTCTCTTAACCTGCCCTGCCCT 3665
QY 3587 ggcctccctgacgtaactaaaccttaaccttattgtgatgagcgttaaggagtagc 3646
Db 3666 GGGATTGTCCAGACTTACTAAGCCCTTCGAACTTTTGTGAGAGAAACAGGCGCTACGC 3725
QY 3647 cggagaggtttaaaccacaaccttagacataggagagagcagctgtgtcactatgcata 3706
Db 3726 CAAAGTGTCTTAACCCAAAACTGGGGCTTGGCTGGCGGCTGCTACTGTGCCAA 3785
QY 3707 gaaacttgatccctgtagcagtggttggccggtatgtcgaagcctatgcagctgtgac 3766
Db 3786 AAGCTAGAACCGTGTGGAGCTGGGTGGCCCTTGTCCATGCGATGTGACGACATGCC 3845
QY 3767 catactgtcaagagcgttgacaatttgacttggagacgaatataactgtataagccc 3826
Db 3846 CGTTCTGACCAAAAGAGCGTGGCAAGCTCACATGGGACAGGCACTGTCATTCTGCCCC 3905
QY 3827 ccatgatgtgagaacatcgttcggagcccccagaacagatgatacgaagcccgagat 3886
Db 3906 CCATGCAGTAGAGGCACTGATTAAAGCAACCCCTGTGCTGCTCTCCAGACGCCGAAT 3965
QY 3887 gaccacatacaaagcgtgcttct---caagaagaggtcacttctgcaccacagccgc 3943
Db 3966 GACCCACTACCAAGCTCTGTGTTCTGACACGAGACGATCCAGTTGGACCAATAGTGGC 4025
QY 3944 tctcaacctgcacactctctgcctgaagaagactgtatgaacagatgactcatgtatgca 4003
Db 4026 CCTAAACCCAGCTACCTCTCTCCCTTA---CCTGAGGAGGGGCTGCCAAGTACAGCTCT 4082
QY 4004 tcaactatgattgagagagcagtggttcgcaagagccttaccagatacaccgctgactg 4063
Db 4083 TGACATCTTGCTGGAAGCCACGGAAGTCTTACGAGCAAGCTCTCTCCAGA 4142
QY 4064 agaagtgctlaacctgttcaactgaagcagctatgtgtgtaaggttaagaggtgac 4123
Db 4143 CGCTGACCAACACTGTGTACAGATGGAGCAGCTTCTGCAAGAGGGGACGCAAGGC 4202
QY 4124 tggggcgagcagtggtgagcgggacccgacgacttgggcccagcagcctgcccgaagaa 4183
Db 4203 CGGAGCAGAGTAAACCAACGAGACCGAGGTAGTGTGGGCCAAAGCACTGCCAGCGGGAC 4262

QY	4184	ttcagcgcacaaaggctcagctcgtcgtccacgcacgaagcttgcggtcgtccgaaggaa	424
DB	4263	atcgccccaagagctgattgattagcgctccaccgaagcctttaaatatggcagaaagtr	4322
QY	4244	atccataacattatcgcagcagcaggtatgcctttgcagctgcacgtaacagggc	4303
DB	4323	gaagctgaattttttacccgattmccgtttatgcttttggccactgcccattttcacgagaa	4382
QY	4304	catcctaaacaagggtgttgccttactcagcagggagggaaatcaagaacaagaga	4365
DB	4383	atatatattgaagcgccgggttgctcacatcagaagaaaaaanaattataaagacga	4424
QY	4364	aattctaaagcctattagaagcctbaacttgcacaaaggctagctattatacactgtcc	4423
DB	4443	gattttggcccttactgaaaggctcttcttccgcccataaagacttagcatattttggccc	4502
QY	4424	tgagcatcagaagaacccaagatcctaactatctagaaggaaacagatgctgcacgggttgc	4482
DB	4503	gggacattcagaaggaaacccgcggagggcaaggggcaacagatggccgaccaaagcggc	4562
QY	4484	caagcagcagcccccagcgtcttaaccttgcctataatagaaacgcccacccaga	4534
DB	4563	ccgagaattaccctcgtgagaaactccagagacttccacacttttgatgaaatttcagc	4622
QY	4544	accg-agcgcagtagacacctagaagacttgcacagagataaaaaagatagaccagtct	4602
DB	4623	ccccattacttcataacatttttcactatgcgtgacatataaagaatcactataact	4682
QY	4603	ctgagcactccggaggg-----gacgcgtctactctatcttggaaggaaatctctcc	4654
DB	4683	aggggccacttattatgcattgcataaagaaagtgttggctttttcagggaaagcctgtatggc	4722
QY	4655	cccaaaagaaggtttagaatactgccaagaatacagctcactaacccactlag-----	4708
DB	4743	tgatcattatcacttcttgaaactatttagattttcttcatcaattgacccaactcagtttctc	4802
QY	4709	aactaaacctgtagaagacttggctcagaacatccctctatctatgtctcgtaggtctacag	4788
DB	4803	aaaaacaaaggctctttctagaaaggaactactgctcttatttcattcgtgaacccggatcg	4862
QY	4769	agtgctgcagctcgtgtgtcgaacatctgtgctccctgcagctgcgttaatgtcactcttc	4828
DB	4863	aacgctcaagacatcactgagacttggccaaagccttgcacagtcattgcacgaactc	4922
QY	4829	cagaaatccccaaggaaagactaaagggggaagccacccagcgctcactgggaagtga	4888
DB	4923	tgccgctcaaacaaaggactgtgaggtttgagggccaccgacccgacccactgggaatttga	4982
QY	4889	cttcaactgaagtaagccggtctaactaaggaacaatactcattgtgttttggtaaac	4948
DB	4983	ttttactcagagtaaaaccttgccctgtrattgggtrttaattctttaaagttttcatatgacac	5042
QY	4949	ctttcagatbvggttagaggtctatccctaactaagaagaagacttcaaccgtgtgtgtctaa	5008
DB	5043	ttttctctgaaagggttagaactttcccaacaaagaaagaaacttgcacaaagtgttaaccaa	5102
QY	5009	gaataactctbvgagaaattttcccaagatttgaataacttaagtgtaataaggtcaagaa	5068
DB	5103	gaaccttacttgaagaaattcttccccagatttggcattccacaggtatttgggaacccgacaa	5162
QY	5069	tggtccagccttctcgtttgcccaggttaagttaaggaagctgcgcagaatcttggagattgtg	5128
DB	5163	tgggcctgcgcttgcttccaaggttaagtcagacagttacgttgaagaaataaataatagac	5222
QY	5129	gaactcgtactgtgtcatcacagaccccaagcttcaggaacaggttagagagatgaatagaa	5188
DB	5223	gaactcgtactgtgtcatcacagaccccaaggttccaggttgaagaaataaataatagac	5282
QY	5189	cattaagagacaccttaccacatctggcacaagaagctgtcatatagttatgtagtgcct	5248
DB	5283	aattcagagacattttacttaatttgacgcttgcacacttgctttagggactgggtgcttct	5342

QY	5249	ccctgcaccttgcgtcttttaagggtgaagaacaccccttgcagatttggccttgaaccccta	5300
Db	5343	gcttccctcagaccctgtattgcagccggcgaacacccggcccccacatgctctacccctata	5402
QY	5309	tgaattgcctcagcggggagcccccccgcttgcagaattgccttgcacatagtcgtga	5366
Db	5403	tgaamaatttatattggcggcaccccccccctgttgaacttccctgcagatccgacatggcAAA	5462
QY	5369	tgtgcctgtcttcccaagcgccttgccttgccttaagtcacaaggcgcctcgaattgggtgaagcagc	5428
Db	5463	ggttacctatataccctctcttcccaagccattttacagcacttcttacctgtgtccacacaga	5522
QY	5429	agcgttgcagaagcagctccgcggagacctactaagga-----ggaagcttgcatttccaca	5482
Db	5523	actctgcgagaccggttggcggcagcctttaccAAAGAACAACTGAGCCGCGTGTGTCCTCA	5582
QY	5483	tcgcttccaagtttgagattcagctctgtttaaagcacaacgcttgcagaagaaacctcgaac	5542
Db	5583	cccttttccgagctgcgtgacacatgagtggtgggtccacagacacacaaacttgaataattctgaAAC	5642
QY	5543	tcggttgcagaagacctactctcgcacttcttgacacacaacagcgtctgaagtcogaag	5602
Db	5643	ccgctgcgaagaagaccttattaccgtcttactgacttacccccacccgctctcaagtgtgacgg	5702
QY	5603	aatcccaacctgatalcatgatcatccacgttlaagccggc-----gcaccc	5662
Db	5703	catttcacacgctgagatccacgctgcacgaaagagctgcacacacagatgtgacccacac	5762
QY	5648	tccgcagctcgggggttggagagcgcgaagaagctgaaattcccttaagcttgccttccatcgc	5707
Db	5763	atccgaattgcacatggcgtgttcaacgctctcaaaattcccttgaagatnagattgataccgg	5822
QY	5708	ccgtgttcccttactactcaacataactcccaagcgcagtagtlaaacgcctctatagaagct	5767
Db	5823	ccggacacctctaatcccttattctcttctgtcttctcaaaaggccacagatccgcacaga	5882
QY	5766	cgaaccccca-----tagaccttattcccttactctgcgtatattgaacctgatacggct	5823
Db	5883	cccgccctccacccctccacacagcttccacatttacctgggaagtgacaaatggggatgg	5942
QY	5824	gtcactgtlaaataagcactcgaagtggtgtccctcagaagcactgttgyccttgaac-----	5878
Db	5943	gagagagatnagggaatnatacgggaacacacccctctgtgacacttgggtggcgaactctccac	6002
QY	5879	-----tgaatttctgcctccgaltgataacccgcgtctaaagcacacctcca	5929
Db	6003	ccagatttggttattgtattagctcttcagtgggcccccctccactgggggctgtagattatcagccc	6062
QY	5930	acctgaatccagtagtataatgtgttc-----tatgtgtccagagcagaagaagaagaataac	5985
Db	6063	ccctatttctctggccccccggggcccccctgttgctcagggagacagcgggaactgttcagacc	6122
QY	5986	tgttggggg-----tcttggggaaactcttctgtga	6016
Db	6123	tgtgcacaagacactcaacagacccttgactccctccacccctgggtgacaacactgtccgg	6182
QY	6017	gatgagagctggtactactccaagatgagagactgcgaatagccgactctctctccaagacc	6076
Db	6183	aacagactttaaagctgcgacacaggttactcattaaatcaatgaaggatttttatgtctgcccc	6242
QY	6077	gggtlaaanaattctccttgttcaatttcgcgcgcgcgcgaagtaacaaatga tgaact-----	6131
Db	6243	gggtacatctgcgcccccggaagcagctctgtgggggttccagactccttcttactgtgcc	6302
QY	6132	-----ataaagaataagagctgtcccca tccgaacta	6165
Db	6303	tcttggggcttcgcagacacccggttaagattactggaagccctctcttcttgggagctac	6362
QY	6166	gattatctaaagaataagtttctacttgaagaagaaacgcggaataattcgaagatgtgataaa	6225
Db	6363	atccagatgacacaaatcttcacactcttaaccagcgtgtttcaggttgcAAAAGACAATTAG	6422
QY	6226	tgtatagactcgtgggaatgatttttatatatatgycgcgggggcaggggtccaacttcaac	6285

Db	6423	TGGTGCAATCCCTTGGCTATCCGGTTTACAAAGCGCGGAAACAGGTCACTCATGGACA	6482
QY	6286	a-----tgccttaagataagacgggacagaacccccgtgycaatg	6330
Db	6483	ACTGACACATATTGGGGCTTACGCTCTTTATGTCCTGGACAGACCCGAGGCGTTACTTTC	6542
QY	6331	ggaccgataaagtlactgctgtacgaggggccccggcccttgagccacgcataactg	6330
Db	6543	GGGATCCGACACTGATTATCAAAATCTTAGGACCTCGGATGCCAATAGGACCAAAACCCGTC	6602
QY	6391	ccgattgcccaatlaaacctgcgtgcgcctgcatacaacagccgcctlagacaacagttacc	6450
Db	6603	CTGGCAGACCAACTTGTTGGTTCCGGCTACTACTAATCCCTTACCAAACTGCGCAAGTCTCC	6662
QY	6451	actgattgatctccatacaacacgcctagaaactccccaggtgttcctgtlaag-----	6504
Db	6663	CCCGCCCTCTATGTTGACTCCCACTTGAATTTCCCGCTCCCACTCCACATCAGGCCCGC	6722
QY	6505	-----aagagacagagactctcagttcatccatccagggagcttccaagccataac	6555
Db	6723	CCAGCAGGAACGGGAGACAATTACTTAATCTTAGTACAGGGAGCTTACCGACTCATAC	6782
QY	6556	tccaccgacctgtagtcacctctctctgttgcttgctcactccaaagggcctccat	6615
Db	6783	CTTACCAACCTGATTAATACTCAAGAGTGTGTTATGCTTAGTGTGTGGACCCCCCTAT	6842
QY	6616	tatgagggatggtcctaaagaagaataatcaatlygaccaaagagcatagaaatcaatlyt	6675
Db	6843	TACGAGGGGGTGGCGTCTAGTACTATTATTCACACCACTACTCTGCCCCAGCTATACGC	6902
QY	6676	acatggyggtcccgaaataaagcttaccctccactgaaagttccggaagggagacatgcat	6735
Db	6903	TCCCTGGCCCTCCCAACACAGGCTGACCCCTCGGAAGTACGAGCGGGAGCTCGCATTA	6962
QY	6736	ggaagaagctcccatcccaacacacacttgctctagactgigtgttaagcagagcc	6795
Db	6963	GGAACTATCCCAAAACTACACAGGCCCTGTGCACACTACCT-----TAAGCAAGC	7016
QY	6796	tcgaaaaatagatlatagtaacctgtttaaocaggtggtggtggtacatgcaatctggttla	6855
Db	7017	AAAGGCTTACTATCTAGTTAGTTGGCCCCACAGGAACTAATGTGGCATGTATACACTGGAATC	7076
QY	6856	acccccgtgttccacactcagctccttaaaccaatcccaagaattctgtgcatggtccaa	6915
Db	7077	ACTTCATGCTTACTGTGCAACCGTCTTATTCGACACACTGATTTGGTGTCTCGTGGA	7136
QY	6916	atgctcccccaggtgtaactacaactcctggggaagtgctcctgtaagaatagcatatgag	6975
Db	7137	TTATGGCCCAAGGTCACCTTACATCTCCACTTACGCTTATAGCCCAATTTGAAAATCC	7196
QY	6976	tataaccagcaaaagaagaccgcatatcccttaacctagctgtaatgctcggaattag	7035
Db	7197	CATA---GACATTAAGAAGACACAGTGTCTTAACTTGCGCTTATTTAGTGGTGGCTA	7253
QY	7036	acggcggtggtcgtlaggaacagggacagctgcctccatcacaggaaccaagcagctag	7095
Db	7254	ACTATGGGTGGCTATGGCGCGGGAGTAAAGGACAGGAATACGC---CCCTGATGGCCAC	7310
QY	7096	aaagagacttggtagcttaacatgcgcgcatagcaagaagaatctccgagccttaagagatct	7155
Db	7311	CAGCAGTTTACAGAGCTCCATGCTGCCGTACAAAGATGATCTCAAAAGAAAGTGAAGAACGCA	7370
QY	7156	gttagcagcctagaagagctccgagcttcttgcttgaagtgtgtctagaacacggag	7215
Db	7371	ATTACTTAACCTAAGAAAGTCTCTTACTTGCTGTGCTGAAGTTGTACTCCAAATTCAGCA	7430
QY	7216	ggaatagatctgctgttctcaagagaagtggtgtlatagtgcagccttlaaagaagaatgt	7275
Db	7431	GGCCTAGACCTGTTGTCTTAAAGAGGAGAGCTGTGCTGCCCTTAAAGAAAGTAATCT	7490
QY	7276	tgctctcatgtatgaataactcaggagccatcagagactccatgaacagcttagaaaaag	7355

Db	7491	TGTTTCATGCTGCATACAGGCGTAGTAGAGATGATGATGAGCCAAATTAAGAGAGAGA	7550
Qy	7336	ttagaagagcgctogaagaagaagaagcgtgcacaggggttggtttgaagaatggttcaac	7395
Db	7551	CTCTCTCAGAGACMAAAACATATTTTGGATGTCAGAGCAAGATGTTGCGAAGATGTTTAA	7610
Qy	7396	aggtctcccttgatgataccaccctcgcttcgtcctgaaggggcccctaaagtcctgc	7455
Db	7611	AGATCCCCCGTGTGTACACGTTGATATTCACATCATG66GCTCTCATTTACTCTA	7670
Qy	7456	ctgttacttaacgtltgggaccttaataalaaggtltgttcctgttgtagaagaaga	7515
Db	7671	CTAATTCGTCTTTTGGACCCCTCCATCTTTAATCGATTTGTTCAATTGTGTTAAAGACAGG	7730
Qy	7516	gtgagtgcaatccagatcacatgagtaactaggcaacaatlaca	7556
Db	7731	ATCTCAGTAGTCACAGGCTTTAGTCTGACATCAACAATATACA	7771

```

RESULT      8
US-08-886-642-8
: Sequence 8, Application US/0886642
: Patent No. 5952474
: GENERAL INFORMATION:
: APPLICANT: Pinter, Abraham
: APPLICANT: Kayman, Samuel
: TITLE OF INVENTION: FUSION GLYCOPROTEINS
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 45 Rockefeller Plaza, Suite 2800
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/886,642
: FILING DATE: 01-JUL-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/110,300
: FILING DATE: 20-AUG-1993
: APPLICATION NUMBER: 07/938,100
: FILING DATE: 28-AUG-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Hone, William J
: REGISTRATION NUMBER: 26,739
: REFERENCE/DOCKET NUMBER: 07763/010002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212/765-5070
: TELEFAX: 212/258-2291
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8323 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
US-08-886-642-8

```

	Query Match	21.3%	Score 1747	DB 2	Length 8323
	Best Local Similarity	55.6%	Pred. No. 0		
	Matches 4055	Conservative	0	Mismatches 3005	Indels 238
				Gaps	
Qy	467	tcgcgagcgctcgcagactcttcgcgctctgggaa--gacgcgagacgagctgcgctggt	524		
Db	498	TCGCCCCCCTCGATGATTTTGTCTTGGCTTGGAAACGACGCCGCGCGCGCTCTGT	557		

Db 2706 GGGCATGGGGCTGGCCGTTGGCCAACTCCTGTATACCTCTGAAGGCAACCTCTAC 2765
Qy 2687 accagatcaagtcagacagtcaccccttgagtagaagctcgagaagaaatttggccga 2746
Db 2766 CCCCCTGTCATTAACCAATTAACCCCATGTCTACAGAGGCAAGTGGGATCAACCCCA 2825
Qy 2747 tgttcaagatataccaagaaggaactcctagttcctgtccaatcccttggatactcc 2806
Db 2826 CATACAGAGACTCTGGATCAGGGAATTCTGTGACCTTCAGTCCCTCCCTGGAACAGCC 2885
Qy 2807 cctgataccggttagaagccttggacaatgatatcgacaagtaacagagcttagaga 2866
Db 2886 CCTGCTACCCGTTAAGAAACCGGGGACTATGATTAAGCCCTGTCCAGATCTAGAGA 2945
Qy 2867 ggtcaataaaaggtgcagacataccccaaggtccgaaccccttaaacctcttag 2926
Db 2946 AGTCAGACAGGGGTGGAAACATCCACCCACCGTGCCCAACCTTACACCTCTTTGAG 3005
Qy 2927 cgcctcccgctgaacggaactgltacagatitgagacttaaaagtgcctctctg 2986
Db 3006 CGGGCTCCCAACCGTCCCAAGTGTACTGTGACTTAAGATGCTTTTCTTG 3065
Qy 2987 cctgagatlaaccccaactagccaaacttttgccttgccttgaatgagagatccaggtac 3046
Db 3066 CCTGAGACTCCACCCCACTCACTGTCTCTGCTTTGAGTGGAGATCCAGAT 3125
Qy 3047 ggaagaacccgggagcgtcaccttgaccgagactgcccgaaggttcaagaactcccgac 3106
Db 3126 GGGATCTCAGACATTAACCTGACACAGACTCCGCAAGGTTTCAAAAACAGTCCAC 3185
Qy 3107 catctttagcgaagccctacacaggaagccttggccaacttaagataccaacactcaagt 3166
Db 3186 CTTGTTGATGAAGCCCTGCACAGGACCTGACAGACTTCGGATTCAGACAGACGACCT 3245
Qy 3167 gacctcctccagtcagtgatgagacttcttgccggaagcccaagaagactgct 3226
Db 3246 GATTCGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3305
Qy 3227 aagaagtagaaggaactgactgactgactgactgactgactgactgactgactgactgact 3286
Db 3306 ACAGAGTACGGGGCCCTTTACAAACCTAGGGGACCTCGGATATCGGGCTCGGCCAA 3365
Qy 3287 gaagcccaagatttgcagagaagatlaacacttgggtggttaacagtttgcgggagcgca 3346
Db 3366 GAAGAGCCCAATTTGGCAAAACAGTCAAGATGTGGGATGATCTTCAAAAGAGGATCA 3425
Qy 3347 ggaatgctgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 3406
Db 3426 GAGATGGCTGATGAGGCAAGAAAGAGACTGTGATGGGGAGCTACTCTCGAAGACCC 3485
Qy 3407 caacaagtagagaagtttggggaagcgtgatttgcagactgtagatcccggtt 3466
Db 3486 TCGCAACATTAAGGAGTTCTTAGGGAGCGGAGGCTTCTGCTCTGTGATGATGATGATGAT 3545
Qy 3467 tggagacttgcagagcccaacttaccgcttgaacccaagaaggaaggaaggaaggaagga 3526
Db 3546 TCGCAAAATGGACCCCCCTTGTACCTCTCAACCAAAAGGGGACTGTGATGATGATGATGAT 3605
Qy 3527 tccatgacacagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 3586
Db 3606 CCCAGACCAAGAAAGGCTTACCAAGAGATCAAGAGGCTCTTAACTGAGGCTGAGGCT 3665
Qy 3587 ggcctcccgagagtaacttaaccccttaccctttagtggatgagcgaaggaaggaaggaagga 3646
Db 3666 GGGATTCGACACTTGAATGAGCCCTTCAAACTTTTGTGAGGAGAAACAGAGGCTGAGC 3725
Qy 3647 ccgaggaagtttaacccaaccccttaggaacatgaggaaggaaggaaggaaggaaggaaggaagga 3706
Db 3726 CAAGGATGCTTAACGCAAAACATGGGGCTTGGCGTGGCGGCTGAGCTGATGCA 3785
Qy 3707 gaagcttgaatccctgtagcaggttggccggtatgctgaaggaaggaaggaaggaaggaagga 3766
Db 3786 AAAGCTAAGACCAAGTGGAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3845

Qy 3767 catactgttcaaggaagcgtgacaaatttgaacttgggaacgaatataactgataagccc 3826
Db 3846 CGTTCTGACCAAAAGACGCTGGAGAGTCACTACATGGGACAGCCACTGATGATGATGATGATGAT 3905
Qy 3827 ccatgtatgagaacatcgttccgagcccccagaacgaatgatatgaacagcccgat 3886
Db 3906 CCATGCAATGAGAGCACTGATTAAGCAACCCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3965
Qy 3887 gaacccatcaaacgctctct-----gacagagaggggtcaacttgcctccacacagccgc 3943
Db 3966 GACCCACTTACAGGCTCTGCTTGTGAGACAGGAGCCAGTCCAGTTCGACCAATGATGAG 4025
Qy 3944 tctaacccctgcaactcttctgctgaaggaatgaatgaacagtagactatgattgca 4003
Db 4026 CTTAAACCAAGCTAGAGTGTGCTCTGCA-----CCTGAGAGGGGCTGCAACATGAGCTGCT 4082
Qy 4004 tcaactatgattgagaagacgtgggttcgcaaggaagccttacaagataccgtctag 4063
Db 4083 TGACATCTTGGCTGAGCCCAACCGAACCTAGACAGATCTTACGAGACGCTCTCCAGA 4142
Qy 4064 agaagtgtaactcgttgcctgaaggaagcactatgtgtggaaggaatgaaggaagga 4123
Db 4143 CGTGACCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4202
Qy 4124 tgggagcagtggtggaaggaagcagagatctggtggaaggaagcagcctgcgaaggaac 4183
Db 4203 CGAGCAACCACTTAACACCGAGACCGAGTATGCTGGGCCCAAGACATGCGACCCGGAC 4262
Qy 4184 ttcagcgaacaaagcgtgagcctcagtcagcctgaaggaagccttgcggtggaagga 4243
Db 4263 ATCGGCCCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4322
Qy 4244 atccataacattataaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 4303
Db 4323 GAAGCTGAATGTTTACACCGATAGCGCTTATGCTTTGGCCATCCCATATTCAGGAGA 4382
Qy 4304 catctataaacaaggggttgccttaactgaagcaggaaggaaggaaggaaggaaggaagga 4363
Db 4383 AATATATGAAAGCGCGGGTGTCTCATATCAATCAAGAAAGAAATCAAAATTAAGAGCA 4442
Qy 4364 aattcaagcattatgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 4423
Db 4443 GATCTTGGCCCTTACTGAGAGCTCTCTCTGCGCAAAAGACTTGAATATTCATTCGCC 4502
Qy 4424 tggacatcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 4483
Db 4503 GGGACATCAAGAGGAAACCGCGAGGCAAGGGGCAACAGATGCGCAACCAAGCGGC 4562
Qy 4484 caagcagcagcccgagcgtgtaaccttctgcttaataagaagcccaagcccaag 4543
Db 4563 CCGAGAGATGAGCCACTAGAGAAATCTCAGAGACTTCCACACTTGTGATGAAATTCAG 4622
Qy 4544 accc-agagcagatcacaccctagaagacttgcgaaggaatgaaggaaggaaggaaggaagga 4602
Db 4623 CCCCTATACATGAAACATTTTCACTATACGGTGAATGACATTAAGATCTGACTAACT 4682
Qy 4603 ctgagactccggaggg-----gacctgctataactatgataaggaaggaaggaaggaagga 4654
Db 4683 AGGGCCACTTATGACAGATCAAAAGAGTGTGGTTTATCGAGGAAACCTGTATGTC 4742
Qy 4655 ccacaagaaggggttagaataatgtccaaggaatataaggaaggaaggaaggaaggaaggaagga 4708
Db 4743 TGATCAATTCACCTTGAATATTTAGATTTTCTTCAATTAATGACCACTCAAGTTTCTC 4802
Qy 4709 aactaaacactcagcagcttgcgtgaaggaatcccttctgaaggaaggaaggaaggaaggaagga 4768
Db 4803 AAAAACAAGGCTCTTTAGAAAGGAATCACTCTCTTATATCATCTGAACCGGGATG 4862
Qy 4769 agtgcgtgactcgttgcataacatgttgcacctgcagcgtgtgtatgtatccttc 4828
Db 4863 AACGCTAAGACATCACTGAGACTTTCGAAGCTGTGACAGTCAATGACGACGAAGTC 4922

QY 4829 cagaataccctcaggaagaactaaggagagccaccagcgctcactgggaatgga 4888
DB 4923 TCCCTCAACAGAGGAGTTCAGAGGCGACCGGACCCACCTGGGAATTTGA 4882
QY 4889 ctccactgagtaagccggctaaatcgcgaataatctatgtttttagaac 4948
DB 4983 TTTCACTAGGTAAGTAACTGGCTGTATGGGTAAATATCTTTTGTTCATAGAC 5042
QY 4949 ctltcagatggtagagcttaccactaagaagagagactcaacgttggctaa 5008
DB 5043 TTTCTGGATGGGTAGAGCTTCCACACAGAAAGAACTGCCAAAGTTTGAACAA 5102
QY 5009 gaaatcctggagaataatlttccaaatlttgaaactaactaagtaaggtcagaca 5068
DB 5103 GAACCTACTAGAAAGAAATCTTCCACACTTGGCATGGCATGACAGATTTGGGAACGACAA 5162
QY 5069 tggctcagcttctgcccaggttaagtcagagagcagcagatgttggatgttg 5128
DB 5163 TGGGCTGCTGCTGCTCCAAAGTACAGACAGTACGCGATTTACTGGGGTGTATG 5222
QY 5129 gaaactgcatltgtacatacagaccaccaagctcagacagtagagagatgaagac 5188
DB 5223 GAACCTACTAGTGGTTCAGACCCGACGAGTTGAGTACAGTGAAGAAATGAATAGAC 5282
QY 5189 catlaaagagacccttaaccaatltgacacagagactgcatlaatgtatgtgtct 5248
DB 5283 AATCAAGAGAGCTTTAACTAAATGTGACGCTTGAACCTGGCTTGAAGCTGCTCTCT 5342
QY 5249 cctgccccttgtgctttagaggtgaagacaccccttgacaggttggctgaccccta 5308
DB 5343 GCTTCCCTTACGCTTATCGAGCCGACACAGCCGCGCCCTGCTGCTGCTGCTCTCT 5402
QY 5309 tgaatgctcctacagggagagcccccgcgttggcagaatltgcttgcacatgtctga 5368
DB 5403 TGAATCTTATATGGGGGACCCCGCCCTTATTAATCTTCCCTGATCGACATGGA 5462
QY 5369 tggctgcttccagccttgttctcctaggtcgaagcgttcgagtgagagcgcg 5428
DB 5463 GGTACTCATTAACCCCTCTCTCAAGCCCATTTACAGGACATCACTGCTGCTGCTGCT 5522
QY 5429 agcgtggaagcagctccggagagcctactcagga-----ggagaactgcaagttccaca 5482
DB 5523 AATCTGAGAGACCGTTGGGCGACGCTTACCAAGAAACACTGACCGCGGAGTGGCTCA 5582
QY 5483 tgccttcaaglttgagatltcagltatgttagagcgccagctgaggaactcagagac 5542
DB 5583 CCTTTCGAGTGGGTGACACAGTGGGTCCGACAGACCAAACTAAATCTAGAAC 5642
QY 5543 tgggtggaagagacctaactcgtactttagacacacacagcgctgtgaagtcgaag 5602
DB 5643 CCGCTGGAAGAGACTTATACCGTCTACAGTACACCCACCGCTCAAAAGTGAACG 5702
QY 5603 aatcccaactgatacctatgcatccacgttaagcggc-----gcacac 5647
DB 5703 CATTCGACAGTGGATTCACGCTGCGCACGTAAGAGCTGCGACACAGATTTGAGCAC 5762
QY 5648 tcccgactcgggttgaagagcgaagaagactggaatccacctaaacttcgacctcag 5707
DB 5763 ATCGGAATTCAGATGGGTGTCAACGCTCTCAAAATCCCTTAATAATAGATTGACCG 5822
QY 5708 cctgttccttactctaataactcccaagcagtagtaacgcttataagagact 5767
DB 5823 CGGAGCTCTATCCCTTAAATTTCTCTCTCTCTCAAAAGGGGCGAGTCCGACGA 5882
QY 5768 cgaaccccca-----tagacctatccctactcgtatgtatgtacccctgatacggt 5823
DB 5883 CCGGAGCTCCAGCCCTACACAGGCTTACACATTTACTCTGGGAAGTGAACAAATGGGATGG 5942
QY 5824 gtaacgttaataagcactcggaggtgtgtcctcctagagagcactgttggcctgaac----- 5878
DB 5943 GAACAGTATGGCAATATACAGCAACCTCTGTGTGATTTGGTGGCAGTCTCTAC 6002
QY 5879 -----tgacttctgctcccgatgattaaaccccgctgtttaaagacacactccca 5929

DB 6003 CCAGATTTGTATGTATGTAGTCTCAGTGGGCGCCGACCTGGGGCTAGAGTATCAGCC 6062
QY 5930 accatgctcggagtagttaggtgtc-----tatgtcccaagcaagaagaataac 5985
DB 6063 CCTATTTCTTGGCCCCCGGGGCCCCCTTGTGCTCAGGGAGCAGCGGAACTGTGACGCG 6122
QY 5986 tgggggg-----ttcgggaatcctctgttaga 6016
DB 6123 TGTCCAGAGACTGACAGAGCCCTTGACCTCCCTACCCCTTGCTGACCACTGCTGG 6182
QY 6017 gatgagctgctgtactcccaagatggagactggaatggtccgactcctccagagac 6076
DB 6183 AACGACTTAACTGACACAGTACCTAATTAATCAAGTAGAGGATTTATGTGTGCC 6242
QY 6077 ggttaaatcctccttgttaactccggccggcagagtaaatgtgaact----- 6131
DB 6243 GGGTACATGCGCCCGGGAGGACAGTCTTGTGGGGTCCAGACTCTTCTACTGTGCC 6302
QY 6132 -----ataaagaataagagctgtcccatcagacta 6165
DB 6303 TCTTGGGGCTGCGAGACACCGGTAGATATCTGGAAGCCCTCTCTTGTGGACTAC 6362
QY 6166 gattactaaagaatagcttcaatgaagaagaacaggaataatcaaatgtgataa 6225
DB 6363 ATCAGAGTACAGACATCTCAGCTTACCCGCTGTTCAAGATTCAGAACAAATAG 6422
QY 6226 tggtagagcttgggaatagtttataatataagcggggagagcgggtccactaac 6285
DB 6423 TGTTCATATCCCTTGGCTATCCGTTTCAAAAGCCGGGAAACAGTACCTCATGACA 6482
QY 6286 a-----ttgccttagatagaagaaggagagaccccttggcaatg 6330
DB 6483 ACTGAGACATATTTGGGCTTACGCTTATATCTCTTGACAGACCCAGGGCTTACTTTC 6542
QY 6331 ggaaccgaataagactgtctgaagagggcccgccggcggcagcagcaacttg 6390
DB 6543 GGGATCCGACTCAGTTATCAAAATCTAGAGCTGGATCCCAATAGAACCAACCCGTC 6602
QY 6391 ccgggtcccaatlaactcctgctgctgagcctgacataaacagccgcttagcaatgac 6450
DB 6603 CTGGAGACACCTTCTCTCCGCTACCTAATCCCTACCAACCTGCGACGTGCC 6662
QY 6451 actgattgattccttacaacagccttagaactcccgagtgcttctgttag----- 6504
DB 6663 CCGGCTCTAGTGGACTCCACATGATTTCCCGGCTCCCACTCCCACTAGCCCCG 6722
QY 6505 -----acaagacagagacttcaagtcacatcccaagagacttccaagcactaac 6555
DB 6723 CAGCAGGAAGGAGAGACAGATTACTAATCTAGTACAGGAGCTTACAGGAGACTCAAC 6782
QY 6556 tccacgagccttgaagcacttcttctgtgtgttcttactcctcagggcctcact 6615
DB 6783 CTTCACAACTGTATTAACCTCAAGAGTGTGTGTATGTCTAGTGTGACCCCTAT 6842
QY 6616 tatgaaggtatgtctaaagaagaataatcaatgttgaacaaagagatagaataag 6675
DB 6843 TACAGGGGGGTGGCGCTAGGTACTTATTCACACATTAACCTGCCCCACTACTAC 6902
QY 6676 acatgggggtcccgaaataagcttaccctcaactgaaatltccgggaagggagactaca 6735
DB 6903 TCCGTGGCTCCCAACACAACTGACCTGTGCTCAAGTACAGACGGGACTGTGATA 6962
QY 6736 ggaagactccccaatcccaacaaacttgccttagactgtgtttagagcagggc 6795
DB 6963 GGAAGAGTCCAAACAACTACACAGGCTGTGTCAACACTACTCT-----TAAAGCAGGC 7016
QY 6796 tcgaaataacagatatttagtacctgttataacaggttggcagctgaactcgtgta 6855
DB 7017 AAAGGGTCTTACTATCTAGTGTCCCAACAGAACTATGTGGGATGTAAACACTGACTC 7076
QY 6856 acccctgtgttcaactcagcttcaacaaatccaaagatltcgtgtatgttccaa 6915

Db 1156 GGAGAGTGGGCCCTACAGAGAGAGCCCTGACCCCTTCCCAATGATGCCGCTGCG 1215
Qy 1143 gagcgagacagagatcgatatacgcgtcgcaactatgacctccatgacagg 1202
Db 1216 GGAAGAAAGAACCCCGGCGGATTTCTACTACTGAGGGGTTCCGCTTGGCTG 1275
Qy 1203 ggcgaattgcagccctccagatattggcccttctctcgtcagctctcctaattgaa 1262
Db 1276 GGAGGAATGAGACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1335
Qy 1263 actaacatcccccttctcgaagatcccccaagcctcaagcggttggatgacctc 1322
Db 1336 AATAACAACCCCTCTTCTCGAGAGACCACTAATTTGACAGCTTTGATCGATCGGTT 1395
Qy 1323 atgtctcaccagcctaattggatgatgtgcaacagctgctgcaacatctcaca 1382
Db 1396 CTCTTACTCATCAAGCCCACTTGGGATGACAGCAAGCTATTAGGAGCGTCTGACG 1455
Qy 1383 accgagagagagagatctgttagaggctagaagaaatgttcttgaggcgagagg 1442
Db 1456 GGAGAGAAAGAACGCGAGTCTCTAGAGCCGAAAGCGGTTCAAGGAGAGAGCGA 1515
Qy 1443 cgaccacagcagttgcaaaatgagatgacatggatattccctgacctgccccggttg 1502
Db 1516 CGCCCACTCAGCTGCCCAATGACATTTATGATGCTTTCCCTTGAACGTCGCCAGCTGG 1575
Qy 1503 gactacaacacgctgtagagtagagagagcttgaataatctacgcagctctggtgcg 1552
Db 1576 GACTTCAACACCCCAAGAGGTAGAACCCACTACTCCATTCCTGCTGCTGCTGCG 1635
Qy 1563 ggtctcgaggcgctcaagacgcccactaattgttgtaagtaagaggtgtagag 1622
Db 1636 GGTCTCCAAAGCGGGGCAAGAGCCCAATTTGGCAAGGTAAAGGATTAACCGAG 1695
Qy 1623 ggaacgaagacacccctcgtatattcttgaagagctcagtagaacctcagaagcttc 1682
Db 1696 GGACCTATGAGTCTCCTCTGAGCCTTTTGAAGAGACTCAAGAGAGGCTATCGAGATAC 1755
Qy 1683 accccttctgactcactcagagagcccaagaagcctcagctgagcctcagctatggg 1742
Db 1756 ACTCTTATGACCCCTGAGAGAGCCAGGGGCAAGAACCAATGCGCATGTCATTCATCG 1815
Qy 1743 cagtcgctcgtgatataagaagaactcagagactcgagaaggtttacaggaagcagag 1802
Db 1816 CAGTCCGCCCGGATATCGGCGAAGTTAGAGCGGTTAGAGATTGAGAGATTGAAGACC 1875
Qy 1803 ttaactgactagtagagagagcagagagagtgatatacaagaagagagacagagagag 1862
Db 1876 TTAGAGACTTAGTAGAGAGAGCTGAAGAGATCTTTAATAAGAGAAACCCCGGAGAA 1935
Qy 1863 aaggaacagagaaag 1922
Db 1936 AGAT 1995
Qy 1923 gagaagaattgactaagatcttggccgagtggttgaaggaagagcagagagagaga 1982
Db 1996 GTGCAAGTA 2045
Qy 1983 gagaagagatttaagaanaattagatcagagccctagacagtagagaga-acctgggcaatag 2041
Db 2046 GTTGTGCTACTGTCGTTTACCGGGGAGAGAGAGAGATGACAGGAGAGAGAGAGAGAG 2105
Qy 2042 gaccacactcgaagaagagcagtgctgatttgaagaanaagagacatgaggagaa 2101
Db 2106 GCCCAACTCGACACGACAGTGTGCTACTGCAAAAGAGAGAGAGAGAGAGAGAGAGAG 2165
Qy 2102 ctgcccagaag 2152
Db 2166 TTGCCCCAAG 2225
Qy 2153 taagaattag 2212
Db 2226 AGACGATTAGGAGAGTCAAGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2285

Qy 2213 ggaagggaacacagttgagttcgtgtgatacagagagagagagagagagagagagag 2272
Db 2286 CGGGGGGCAACCCGTCACCTCTTACTGATGATCTGGGGCCCAACACTCGTGTGACCCA 2345
Qy 2273 accattaggaanaacataagaanaaaacccctgggtgagagagagagagagagagagag 2332
Db 2346 AATTCCTGAGACCCCTTAAGTACAGAGCTGCTGGGTCCCAAGGGGCTACTGAGAGAGAG 2405
Qy 2333 gatacagtagactaccagaagacgcttgacttggagtgaggagagagagagagagag 2392
Db 2406 GTATCGCTGAGACCGAGATCGCGAGAGCTAGGCAACCGGATAGAGTCAACCATTTCT 2465
Qy 2393 tctgtgcalccctgagtgccagtagaccctttagttagaagacttaccagaagatgg 2452
Db 2466 CTTCCATGTACAGATTTGCCCTATCTCTGCTAGGAGAGAGATTGCTACTTAACATAA 2525
Qy 2453 agctcaaatcttcttga--acaaagaaagacagagagtgctgtgataaanaaacat 2509
Db 2526 AGCCCAATTTCACTTTGAGGAGATCAGAGAGCTCAGGTTGGAGCAATGGAGAGAGCCCT 2585
Qy 2510 cactgtgtgacccctccaaatagaagaatgataatcttcccaagtaagcc 2569
Db 2586 GCAAGTGTGACCCCTTAAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 2645
Qy 2570 tgataagatatacagctc--tggttgaagagagagagagagagagagagagagag 2626
Db 2646 AGATGTGCTCTAGAGGTCCACATAGGCTCTGTGATTTCCCAAGGCTGGGAGAGAGAG 2705
Qy 2627 aggaatgag 2686
Db 2706 GGGGATAGGGGCTGGCGGCTTGGCAAGCTCTGATGATCAACCTGTAAGAGCAACCTTAC 2765
Qy 2687 accagatatacagacag 2746
Db 2766 CCCGTGTCCATTAACCAATAGCCCACTGTCACAAAGAGCAACTGGGATTAAGAGAGAG 2825
Qy 2747 tgtcaagaataatcaacag 2806
Db 2826 CATACAGAGAGCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2885
Qy 2807 cctgcaacagagtagaag 2866
Db 2886 CCGTCAACCCGTTAAG 2945
Qy 2867 gttcaataaagagtgag 2926
Db 2946 AGTCAACAAACGGGTGGAAGAGATCCACCCAGCCGAGCCCAACCTTACCAACCTTTGAG 3005
Qy 2927 cgcctcccgccgag 2986
Db 3006 CGGGCTCCCACTGTCACAG 3065
Qy 2987 cctgagatlaaccccaactag 3046
Db 3066 CCGAG 3125
Qy 3047 ggaagaacccggag 3106
Db 3126 GGAATCTCAGCAATTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3185
Qy 3107 catctttag 3166
Db 3186 CCGTGTGAG 3245
Qy 3167 gaccctccctcagtag 3226
Db 3246 GATTTCGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3305
Qy 3227 agaaagtagaag 3286
Db 3306 ACAAGTAG 3365

Db 5523 AGTCTGGAGACCGTTGGCGGACGCTTACCAAGAACACTGGACGGCGGTAAGTGCCTCA 5582
QY 5483 tgcctccaaagtggagattcagctatgtatagcccaacgctgcaggaaaccttgagac 5542
Db 5583 CCCCTTCCGAGTGGGTACACAGTGTGGGTCGCGACACCAACAAATCAATTAAGCC 5642
QY 5543 tccgttgaaggagacctatctgtactttgaccacacacagcgctgtgaagtcgaag 5602
Db 5643 CCGGTGGAAGAGACCTTATACCGTCTACTGACTACCCACCGCTCTCAAAATGAGACG 5702
QY 5603 aatccccaacttgatccatgcatcccaagcgttaagccggc-----gcacac 5647
Db 5703 CATTCGACGCGGATTCACCGCTGCCAGTAAAGGCTGCGACACAGCATTTGAGCCACG 5762
QY 5648 tcccgactcgggttggagagccgaagaagactggaaatcccccttaagcttcgctccatcg 5707
Db 5763 ATCGGAATCGACATGCGGTTCACACGCTCTCAAAATCCCTTAAAGTAAAGATTGACCCG 5822
QY 5708 ccttgctcctactactaacaataactccccagccagtaglaaagccttatagacagt 5767
Db 5823 CGGAGACTCTTAATCCCTTAATTTCTTCTGTCTCTCAAAAGGCGCAGATCCGACGCA 5882
QY 5768 cgaaccccca-----tagaccttatacccttaccctgctgataatgaacctgatacgggt 5823
Db 5883 CCGGCTCCAGCCCTCACAGGCTTACACATTAACCTGGGAGATGACCAATGGGGATCGG 5942
QY 5824 gtacctgataatagcactcaggtgtgtcctcctagagagccctgtgtgacctgaac----- 5878
Db 5943 GAGACATATGGCAATATAGGCAACACCCTGTGTGGACTGTGGTGGTGGTGGTGGTGGTGG 6002
QY 5879 -----tgaccttgcctccgattgataaccacccgctgttaaaagacacccctccca 5929
Db 6003 CCATATTTGTGTATGTATGTAGCTCAAGTGGGCGCCGACATGGGGGCTAGAGTATCAGGCC 6062
QY 5930 acctagtcogtagttagttagtgc-----tatgtgcacagacacagagaagaagaataac 5985
Db 6063 CCGTATTTCTCGCCCGGGGGCCCGCTTGTGTGTCAAGGAGCAGCGGGGAACGTTGACGGC 6122
QY 5986 tgttgagg-----tcttgaggaaactcctctctgtaga 6016
Db 6123 TGTCCAGACACTGCACAGACCCCTTACCTCCCTCAGCCCTCGGTGCAACACTGCGCTGG 6182
QY 6017 gatgagctgcgtacactcccaacgaatgagagctgaaatgagccagctctctccagacc 6076
Db 6183 AACAGACTTATAGCTGAGACCAAGTAACTCAATTAATCAAGTAGAGGATTTTATGTGCCCC 6242
QY 6077 gggtaaaaattctccttctgtcaatcccgcccgagcaagtcaaaaatgatagaact----- 6131
Db 6243 GGGTCACATGCGCCCGGGGAAGCCAAAGTCTGTGGGGTCCAGACTCTTCTACTGTGCC 6302
QY 6132 -----atataagataagagctctccccaacagactta 6165
Db 6303 TCTTGGGCTCGAGACACAGCGTAAAGTAACTGGAAGCCCTCTCTTCTTGGGACTTAC 6362
QY 6166 gattactaaagataagttcaactgaaaggaacaggaataatctcaaaaatgataaa 6225
Db 6363 ATCAGATGAGCAACAATCTACCTCAACCTGATGAGTGTGATGATGATGATGATGATGATG 6422
QY 6226 tggatagagctggggaatagttttttatataagcgaggagagagagagagagagagagagag 6285
Db 6423 TGGTGCAATCCCTTGGCTATATCCGCTTACAAACGCGGGGAACAGGTCTCTCTGAGCA 6482
QY 6286 a-----tgcctttagaataagagcggggaagagagagagagagagagagagagagagag 6330
Db 6483 ACTGAGACTATTTGGGGCTAAGTCTTATGTCTGTGAGACGAGCCAGGAGGCTTACTTTTC 6542
QY 6331 ggaacccgaataaagtaactgctgaacaggggcccccgccctggagccacagcataacttg 6380
Db 6543 GGGATCCGACTCAGTTATCAAAATCTAGACCTCGGATCCCAATAGGCAACAAACCCGCTC 6602
QY 6391 ccggtgcccacataaactcgtcggcgtacataacacagcccgctagacaagtaac 6450
Db 6603 CTGGGAGACCAACTTTCCTTCCGCTTACTTATCCCTTACCCCAAACTGCCAAGTCTCC 6662
QY 6451 actgattgatctcaccacaacagccttagaaactccccagggtgtctctgttaa----- 6504
Db 6663 CCGGCTCTAGTTGACATCCCATTTGAATTTCCCGGTCGCCACATCCCATCAGGCCCGG 6722
QY 6505 -----acagacagagactcctcagctcatalcaagagagagagagagagagagagag 6555
Db 6723 CCAGCAGAGACGGGAGACAGTACTTAATCTAGTACAGGAGGCTTACAGGCACTCTCA 6782
QY 6556 tccacagacacccagagacactcctctgtgtggtctgttctatccctcagagccctccat 6615
Db 6783 CTTACCAACCCGTATTAATCTCAAGAGTGTGTATCTCTAGTGTGTGTGTGTGTGTGTGTGT 6842
QY 6616 tatgagggatggtcctaaagaaatcaatgtaacaaagagagagagagagagagagagagag 6675
Db 6843 TACGAGGGGGTGTGGCGTCTAGTACTTATTCACACATTAATCTGCCCCAGCTAACGTC 6902
QY 6676 acatgggggtcccgaaataagcttaacctcaactgaagtttccggagagagagagagag 6735
Db 6903 TCCGTGGCTCCCAACACAGCTGACCTGTCCGAGAGTGTGAGCGGGGACTGTGCTA 6962
QY 6736 ggaagagctcccccaccccaacacacacttgctatagtagtgtgttatagagagcc 6795
Db 6963 GGAACAGTCCAAAAATCTACAGGCGCTGTGCAACACTACCT-----TAAGCAGGC 7016
QY 6796 tcaagaaatcaglatatagtagtctgttatacaaggtgtgtgagagagagagagagagag 6855
Db 7017 AAGGGCTTACTATCTATGCTAGTCCGCCACAGAACATATGTGGGCAATGATGATGATGATG 7076
QY 6856 acccctcgtgttccactcagcttcaacacacacacacacacacacacacacacacacacac 6915
Db 7077 ACTCATGCTATCTGCAACCGTGTCTTAAATCCACACACTGACTATTTGGTGTCTGTGAA 7136
QY 6916 atcgtcccccagatgactacatccctcagagagagagagagagagagagagagagagagag 6975
Db 7137 TTATGGCCAGGGTCAACCTACCATCTCCCACTTACGTATAGGCAAGTAAATATCC 7196
QY 6976 tataaccgacaaaag 7035
Db 7197 CATAT--GACATTAAGAGAACAGACAGTGTCTTACCTTGGCTTATTTAAGTGGGCTA 7253
QY 7036 accgagctgtgtgag 7095
Db 7254 ACTATGGGTGAGATCCCGCGGAGTAGGAGACAGAACTACCG-----CCCTGCTGCCAC 7310
QY 7096 aaagactgtgtgagctatcagcagcagagagagagagagagagagagagagagagagag 7155
Db 7311 CAGCAGTTTCAGAGAGCTCCATGCTGCGTACAAAGATGATCTCAAGAGTCAAGAAAGTCA 7370
QY 7156 gttagcaacctagaag 7215
Db 7371 ATTTACTTACCTTGAAGAAAGTCTTACTTGTGTGAGGTGTGATCTGAGATGAGAGCA 7430
QY 7216 ggaattagctgtgttcttaag 7275
Db 7431 GGCCTTACAGCTGTGTCTTAAAGAGGAGAGAGTGTGCGTCCCTTAAAGAAAGATGT 7490
QY 7276 tgcctatgtagatcaactcag 7335
Db 7491 TGTTCATAGCTACAGCTACAGGCTTAAAGAGATGATGATGATGATGATGATGATGATG 7500
QY 7336 ttagaagagcgttcgaag 7395
Db 7551 CTCTCTCAGAGACAAAACTATTTGATGTGAGAGCAAGATGATGATGATGATGATGATG 7610
QY 7396 aggtccctcgtgagatgagacacccctgtcttctgtcctgaagagagagagagagagagag 7455
Db 7611 AGATCCCGCTGCTTACACAGTGTGATGATGATGATGATGATGATGATGATGATGATG 7670
QY 7456 ctgtacttaacaggtgtgagcctgtcttaataaagttgtgtgcttgtgtgagagagag 7515
Db 7671 CTATTTCTGCTTTTGTGAGCCCTGATTTAATGATGATGATGATGATGATGATGATGATG 7730

QY	4065	gaatgcttaaccctggcttccactggaagcaagctaagtggtgtaagaagtaagatgact	4122
Db	3621	gccgcacccaccttgctatgcggagggagagcttctctgcgaagaaaggcagcgttaaggcc	3562
QY	4125	gggggagcagtgcygagccggagcccgaaagctctggccgaagcctggccgaagaact	4184
Db	3561	ggagcagcgggtgaccacttgagacttgaggtatattggcccaaggccattggccaccgggaca	3502
QY	4185	tcagcgcaaaaagcttgactctatgcccctacgcaagctctgycgttggccgaagga	4244
Db	3501	tcggccccaagagctgaaactgatatggccctaccaccaagccctaaagatggcacaaagtaag	3444
QY	4245	tcacataacattatcagcagcaggtatgcctctggactgacacgtlaccggygc	4304
Db	3441	aagctaaatggtttatctgatatgcctgttagcgtttggccacggcccatattcatgagaa	3382
QY	4305	attctatacaaaaagggttgcttactctagcagggaggaataaagacaaagaagaa	4366
Db	3381	attatracagaagccggggttgctcacatcagaaagaaagatcagaaacagacagag	3322
QY	4365	attctaaagctatagagccttacaatttcgcaaaaaggctagctctataacagtcct	4424
Db	3321	attcttagccctactaaaggctctgttcttcccaaaaagacttaagattattgttggccg	3266
QY	4425	ggacatcagaagcccaaatctcatatctagaaggagaccagatgycgtlaccggttgc	4484
Db	3261	ggacattctaaaggaamaacacgcgacagagccagggccacacggatggccacacacggcc	3202
QY	4485	aagcagcgacggcccgactgta-----accctcgtctataatagaaagcccaagc	4537
Db	3201	cgagtagtagcctactagagaaactccaggaactctcacacttctgatatamaaaactcaac	3142
QY	4538	cccagaagcccaagcagacagatcacccctagaagcctgcgaagatataaagaatagaca	4597
Db	3141	ccctatnaccatgaaacactttctactatnaccgtractgacacaaagattttgaccaaacta	3082
QY	4598	g - -ttctctagactccggaggggagcctgctatacctcatatggaaggaatccctgc	4655
Db	3081	ggagccacttatgacactgacggagaaatattggggtcactcamaagaaagcgtttatggct	3022
QY	4656	cacaaagaagggtlagaatatgctccaaagatatcatcgtctacacccact - agaaactaa	4714
Db	3021	gatcaattcaacctttgagttactgagactttcttccaccaatttaccacactcagcttctca	2962
QY	4715	acacctgcagcagttgtgtcagaacatccacctatactgcttcagagctacaagagtgyc	4774
Db	2961	aaaaaaagcgtctcttccagagaaagcccccagtcctctactacatcctgaacgggatga	2902
QY	4775	tgactcgtgtgccaacatgtgtgcctccgcagcctggttaagttaactcctccaagat	4834
Db	2901	acacttcaaaaatattactctgagactccaaagcttgacamaagtcamaagtcacagcaagct	2842
QY	4835	accctccggaagaagactaagg-----ggaagccaccaagcgcgtcactggygaagtgc	4889
Db	2841	gccctttaaagacagaaacttagggcccgccggccatctggccctggcgcacactggagatgat	2782
QY	4890	ttcactgaagttaagccggctaaatcggcaaaacatactatctgtttttgttgaagcc	4949
Db	2781	tttccagcaggtaaaacttgattgtatggcttttaagtatcttttggttttgttgaatct	2722
QY	4950	tttccagatggtlagaagcttaccactaagaagaagacttcaacggttggtgactaag	5009
Db	2721	ttttcttgctggatagaaagctttcccaacttaagaaagaaacggccaaagtggtgacaaag	2662
QY	5010	aaaatactggaagaaattttccaagaattgtaatacctaaggtataatggctcagaacat	5069
Db	2661	aaactgttaaaagagatcttccctaggttcggccatggccgcagagttttgggaaactgacaa	2602
QY	5070	ggtccagcttgctgcccggaatgtaagagcctgcgcgaagatatgttggtatgttgg	5129
Db	2601	gggcttgccttctgcttccaaagtgatctcagacagtgctggccgattcttggggattgattgg	2542
QY	5130	aaactgattgtcatacagaccccaagctcagaacaggtlagaagatgtaatatagacc	5189

Db	2541	AAATTAACATTGGCTATACAGACCCCAAGCTCAGGTACAGTAAGAAATGAATGAGACC	2482
QY	5190	attaaagagacccttaccaaattgacacagagacgtgacatratatgatgtgtctc	5249
Db	2481	ATCAAGAGAGACTTTTAACTAAATTAAGCCTTGCAACTGGCTCTAAGAACTGGGTGCTCTCA	2422
QY	5250	ctgccccttltgtcttllttaaggttgaagaaaccccttgaagagtlitgtgtgacccctat	5309
Db	2421	CTCCCTTAGGCGCTGTACCGAGCCCGCAACGCGGGGCCCGCATGGCTCACGCCATAT	2365
QY	5310	gaattgtctctacggyggagcccccccgcttggcgaaattgtgcttltgacatagtgtcat	5365
Db	2361	GGAATCTTAATATGGGGCACCCCGCCCTTGTAACCTTCCCTGACCCCTGACATGACAGA	2302
QY	5370	gtgcgtcttcccgaccttgttctctaggtctaaagggcgtctgaatgtgtgtgaggaaga	5423
Db	2301	GTTTACTTAACAGCCCTCTCTCCAAAGCTACCTTAAGGCCCTTCAAGCACTACAAAGAGAG	2242
QY	5430	gcgttgaagcagctctccggaagacctactacagagagagacttga-----agtlccacat	5485
Db	2241	GTCTTGAAGCCACTGGCTGGGGCTATACAGACCAAGCAAGACGACGACTGATFACACAC	2182
QY	5484	gcgttccaaglttggagattcagltcgtatgttagagcccaacgaagcgtgttgaagtgaaga	5603
Db	2181	CCCTTCCTGTCGGCGACACCGTGTGGGTACGCGGACACAGCATGAAGAACTTGGAACTT	2122
QY	5544	cgttgaagaggaacttatctcgtacttltgtaccacaacgaagcgtgttgaagtgaaga	5603
Db	2121	CTTTGGAAAGAACCTTATACGCTCTGCTGTACCAACCCCAACGCTCTCAAACTTAACAGCGC	2062
QY	5604	atccccacccttgatcatalgacatcccaagtttaagcggcgccacctcccgactcggggtg	5653
Db	2061	ATTCGTGGTGGATTCACAGCGGCTACAGTAAAGCGGCAACACCTCTCCGGCCGGAACA	2002
QY	5664	agaagccgaanaagacttgaatcccttaagcttgcgctccatcgctgttltcttactct	5723
Db	2001	GCATCAGAGACGACATGGAAGGTCC---ACGTTTCAAAAACCCCTTAAGTAAGATTA	1945
QY	5724	aacataactctcccaagccagltagtaagcctctatagacagctcgaaaccccatagac	5783
Db	1944	ACCCCTGGGGCCCCCTATAGTCTCTGGGATCTTTAATAAGGGCAGAGATGATGATGACAC	1885
QY	5784	tt--talccttactctgctgatatttgaccttgatacgggtgctacgttgaataagact	5841
Db	1884	ATGACAGCCCTCACACAGGTCTTCAAACTTATCTTGAAGATTTACCAACTTAATGACAGAC	1825
QY	5842	cgaagttgtctctccttagaagcacactgttgagccttgaacatcttctgcctcgatgtt	5901
Db	1824	AAACAGTAACTGACTACCTCCCTCTGGGGCAATGACAAATGCCCTTCTGATGCTACT	1765
QY	5902	aaccgcgtgtlaaagcacactcccaactagtcgtcgtatgtatgtgtctatctgtcgc	5961
Db	1764	TGCAATTTGTGGATTTTAAATGAGGGAGAGATTTGGAGTATGACCTGACTTGGGTGTGCACTC	1705
QY	5962	ccaagcaacaggaagaagaataactctgtggggttctcgggaatcc-----ttctgtgag	6014
Db	1704	CCGGGGAGAGAAACGGGGCAAGACATTTGACTTCTATGTTTGGCCCGGGCATATCTGTAC	1645
QY	6015	gagatggagctgtgtcaaccccaacagatggagacttgaagaatgtgcgactctctctca---	6071
Db	1644	CAACAGGGGTGAGGGGCGAGAGAGGGGTACTGTGCAAAATGGGGGTGTGAGACCACTG	1585
QY	6072	ggaccggttataaattctctcttltgcaattccggtccgggcaagtaacaaatgatgaaact	6131
Db	1584	GACAGCATATCTGGAAACCATTCATCATTCATATGGGACTTAATTTCCCTTAAGCGAGGAACA	1525
QY	6132	ataataagataaagcgtgcctcccatcagaactagatatataaag-----ataagtcca	6187
Db	1524	CCCTCGGAATCAAGGCCCTGTTATGATTTCTCAAGTGTCTCCAGTGGCATCAAGGGGTG	1465
QY	6188	ctgaagaggaacacaggaanaattacaaagtgtgataaattgttatbaagcttggagatagt	6247

Db 1464 CCACACCGGGGNGATGATCAATCCCTAGTCTAGANTTCATGACGGGATAAAG 1405
Qy 6248 ttatatataggggggagcgggtccacttaaccatcgcttaggataagagc 6307
Db 1404 CCAGCTGGATGGCCCCAAAGTATGG-GGACTAAGACTGTAACAATCCACGGGATGAC 1346
Qy 6308 ggaacacccccctgtgcaatggagccgaataagtaactgctgaaacagggccccc 6367
Db 1345 CCGGTGACCGGGTTCTCTTTACCCCGCAGGTCTCTCAATATAGGGCCCCGATCCCAT 1286
Qy 6368 ccttggagacacacataactggc-ggtgcccaataaactgctgctggcctgacata 6426
Db 1285 GGGCTTAATCCCGTGAATCATCCCACTACCCCTCCGACCCCGCATCATGAGCTC 1226
Qy 6427 aacacgacgcttagaacaacagtaacacgtaactgattcttaccacacagccctagaatcc 6486
Db 1225 CCGAGGCTCTCTAGACTCTCTTACAGGGCGCAGGCTTATAGTTCCTGGGACTGCCCCA 1166
Qy 6487 caaggtgtcctgttaagacagagagactcttcagttccatccagggagcttccaa 6546
Db 1165 CCGTCTCAACAACCTGGGAGGAGACAGGCTGCTAAACCTGTAGATGAGATACCAA 1106
Qy 6547 gcaatcaactccacacgacccctgtagcactctctctgtgtgtgtgtcttctccag 6606
Db 1105 GCACCTCAACCTCACAGTCTGACAAAACCCAGAGTGTGTTGTGTGTGTGTGTGTGGA 1046
Qy 6607 cctccttattagggagtagctaaagaagaataatcaatgtgacacaaagacataga 6666
Db 1045 CCCCCCTACTACGAAGGGGTTGCCGTCTAGTACTTACTTCAACCATCTCTGCCCCA 986
Qy 6667 aatcaatgtataggggttccgaataaacttaaccctcactgaagttccgggagag 6726
Db 985 GCTAAGTCTGCTGGGCTCCCAACACAGCTGACCTGCTCGAAGTAAACGGAGACAGA 926
Qy 6727 acatgcataggaagaagctcccccacatccacacacaccttgcataagtaactgtgttat 6786
Db 925 CTCTGCTAGGAGAGTGTCCCAAAACCATATAGGCTGTATATACACCC-----AA 872
Qy 6787 gacagagcctcagaataatcattatgacgtgtlaacaggtggtgagcagcaat 6846
Db 871 AAGAGAGAGAGAGGGTCTCTACTATCTGTGCTCTCCCGGGGACCATTTGGGGTTGCAAC 812
Qy 6847 actggtttaaaccctgtgttccactcactgacttcaacacaaagatttctgtgtc 6906
Db 811 ACCGGGCTCAGCTCCCTGCTCTACTACTACTACTACTACTACTACTACTACTACTACTACT 752
Qy 6907 atgtgcacatcgctcccccaggtgtaacacatccctgaggaagtgctgtgacgaat 6966
Db 751 TTAGTTGAACTGTGGCCCAAGATATTATACACTCCCGCATTTATATGTATGTACAGCTT 692
Qy 6967 gactatcggtataacgacaaagaagacccglatcccttaccctagctgtatgtctc 7026
Db 691 GAACGAGC---TACCAATATATAAAGAGAGCAGATATGACCTGGCCCTTCTACTA 635
Qy 7027 gattatggagagccgtgtgctgaggaagagagagagctgctgacatcagaagacag 7086
Db 634 GGAGGATTAACCATGGGAGGATTCAGACTGAATAGGAGAGGGGACACTGCTTAAAT 575
Qy 7087 cagctagagagagagctgtgagctacatcgccatgagagagagagagagagagagagag 7146
Db 574 AAAACCCAGCA---GTTTGGACAGCTTATGCTGCTATTCAGACAGACCTTAAAGAGAGTC 518
Qy 7147 aagagctgtttagcaactaagagagctcctgactcttctgtcgaagtggttctacag 7206
Db 517 GAAAGTCAATTAACCACTGAAGAAAGTACACTGCTGTCTGTAAGTGTCTCTACAG 458
Qy 7207 aacggagggagtagatctgtcttctaaagaaggtgggtatgtgacgcttaaa 7266
Db 457 AACCCAGAGAGGCTGATTTGCTATCTCTAAAGAGGAGGAGCTCTGCGCAGCCCTAAA 398
Qy 7267 gaagatgttcttctatgacatcagagagagagagagagagagagagagagagagagag 7326
Db 397 GAAGATGTTGTTTATGACAGCCACAGGGGCTAGTGAAGAGACAGCATGCCCCAATTA 338

Qy 7327 agaaaaaagttagagagcgctcgaagggagaaagagctgacacaggggtgttgaagga 7386
Db 337 AGAGAAAGGCTTAATACAGACAAAAAATATTGAGACAGCGCAAGATGTTCCAGAG 278
Qy 7387 tggltcaacagctcctcttgatgacacacccctgcttctgtcttgaagggccctagta 7446
Db 277 CTGTTAATATGATCCCGTGGTTTACCACTTAATCTCCACCATCATGAGGAGCTTAATA 218
Qy 7447 gtctgtcctgttacttaacagttggccttcttataatagttgttgccttgt 7506
Db 217 GTACTCTTACTGATCTTACTCTTTGGACCTTGATTTCTCAATGATTAATCCAAATTTGT 158
Qy 7507 agagacagagtagtgcagtcacatcagtaactgacttaggcaagatacca 7556
Db 157 AAGACAGGATATCATGATGTCACAGGCTAGTATTGACTACCAATATCA 108

RESULT 11
US-08-850-961-1
; Sequence 1, Application US/08850961
; Patent No. 6013517
; GENERAL INFORMATION:
; APPLICANT: Respass, James G.
; APPLICANT: De Polo, Nicholas J.
; APPLICANT: Chada, Sunil
; APPLICANT: Sauter, Sybille
; APPLICANT: Bodner, Mordechai
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Chiron Corporation, Intellectual Property - R440
; STREET: P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,961
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kruse, No. 6013517man J.
; REGISTRATION NUMBER: 35,235
; REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3520
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-850-961-1
Query Match 20.6%; Score 1687.4; DB 3; Length 8332;
Best Local Similarity 55.4%; Pred. No. 0;
Matches 3964; Conservative 0; Mismatches 2981; Indels 208; Gaps 28;
Qy 585 aatatggagacagatgacataccccccttagttgactctgacacatggagctgaagt 644
Db 618 AATATGGGACAGACTGTACCATCTCCCTTAAGATTGACCTTAGGTGACATGGAAGATGTC 677
Qy 645 agatcagagctatataattgtcagttcaggttaaaagaagcccttgacagactctgtc 704

D	678	GAAGGATGCGCTCACAAACCAATCGGTAGATGTCAAGAAAGACGTTGGTTACCTTCTGCG	737
O	705	gactctgaatgagcaacaactcgatglttggatctgagccatcagaggggaaccttaattctgaa	764
D	738	TCTGAGAAATGGCCAACTTTAAAGTGGATGGCCGGAGACGGCACTTTAAACGAGAC	797
O	765	attaccctgagctttaagagcaatcatlittcagacttggaccggctctatcctgtcag	824
D	798	CTCATCACCCAGGTTTAGATTCAGAGTCTTTTCACCTGGCCCCGATGAGACCCACGACG	857
O	825	gagccctatccttctagcttggcaagatcttggcagaagaatctccgcacatggtttaaaca	884
D	858	GTCCCTCATATCGTAGACTGGGAAGCCTTGACTTTTGACCCCGCTCCCTGGGTCAAGCC	917
O	885	T-----gctcaataaacaagaagccaggtcccgaaacctgcgc	925
D	918	TTTGTACACCTTAAGCCTGCGCCCTCTTTCCTTCACATCCGCCCCGCTCTCCCCCTTGA	977
O	926	tcttggagagaaacaacaacactgcgcgaaaaagtcagagccctctctctgatatcac	985
D	978	CCTCCTGTTTGACCCCGCCCTCATCTCTCCCTTATCCACCTTCACCTCTTCTAGGC	1033
O	986	ccgagatcagagagcgcgcaacttggccggaacccaacacctgtctcccaaccccttacc	1042
D	1038	GCCAAACCTTAACCTCAAGTTCTTTTTCAGACAGGGGGGGCGGCTCATACCACTTACA	1099
O	1046	cagacagaggtgctgcttgaagggaacctctgcctctcttggagctcgggttggaggacc	1107
D	1098	GAACACCCCCCGCTTATAGGAGAACCAAGACACCCCTTCGACAGGAGGAGAAATGCT	1157
O	1106	tgctcccgagactggagccggagagcgacaccccgagcg-----	1147
D	1158	GGAGAACGAGACCCCTGCGGAGAGGACCGGACCCCTCCCATGAGCATCTGCGCTACGT	1211
O	1148	gacagcagagatcgcatataccgtgc-----gacctatgcccctccaatctgcagag	1207
D	1218	GGGAGACGGGAGCCCTGTGGCGCATCCATCACTCCGAGCATTCCTCCCGCGCA	1277
O	1203	ggccaacttgcagccccccaagtattgccccttctcttgcagatctcataattggaaa	1267
D	1278	GGAGGAAGAGACGCTTCAATCACTGCGCTTCTCCTTTCGACCTTTACAACTGMAA	1337
O	1263	actaacatcccccttctctcgagagatcccccaacgcctcaggggttgttggatccct	1322
D	1338	AATATATACCTCTCTTTTCTGAAAGATCCAGTAACTGACAGCTCTGATCGACGTGTT	1397
O	1323	atgtctctaccagccttcttggatgtattgcaacagcctctcagaacactctca	1387
D	1398	CTCATCACCCATGACGCCACCTGGAGACATGTCACACTGTTGGGACTCTGCTGAC	1457
O	1383	accgagagcagagagaaattctgttaagggtcagaaaaaatttctcttggggccgacgg	1442
D	1458	GGAAGAAAGAAACACAGGGGTGCTTTAGAGGCTAGAAAGCGGTGGGGCGATGATGG	1517
O	1443	cgagaccagcagtttgcgaatgtatgcaatgagatcttcccttgaacttgcgccggttg	1502
D	1518	CGCCCCACTCAATGCCCATATGAGATGCGCGCTTTTCCCTCGAGGGGCCAACAATGG	1577
O	1503	gactacaacacgcttgaagtttagagagagcttgaaaatcatatgcgcagagctctgtggc	1562
D	1578	GATTATCACCCACGAGGAGTAGAAGCAACCTAGTCCACTATTCGGCAGTTGCTCTACG	1637
O	1563	ggtctccggggcgctccaagacggcccactaatlttggctaaagttaagagatgtacag	1622
D	1638	GGGTTCAAAAGCGGGACGAAACCCCACTAATTTGGCCAAAGTAAAGAAATAACACAA	1697
O	1623	ggaccgagaaactccctcgtgattcttcttggagagctcatgaaagcctcagcggctc	1682
D	1698	GGGGCCAAATGATCTCCCTCGGCCCTTCCTAAGAGAGACTTAAAGAAAGCTATCGCAGTAC	1757
O	1683	accoccttgaactcactcagagggcccaagaagcctcagtgagccctggccttaattgg	1742
D	1758	ACTGCTTATGACCTTAGGACCCAGGGCAAGAAACTAATAGTCTCTATGCTTTCAATTTGG	1817

QY	1743	caagtcgctctcgtatcatalccaggaagaaccticagagaclygaaggtltacaggaagctcg	1802
Db	1818	CAGTCTGCCCCAGACATTTGGGAGAAAGTTTGAGAGAGTTTGAGAGATTTTAAACAAAGACG	
QY	1803	taagctgctcagtagagagagcaggaagaaggttatcagaagaaggagacagaagaagag	1865
Db	1878	CTTGGAGATTGTTGGTTAGAGAGCGAGAAAGAAATCTTTTAATTAACGAGAAACCCCGGAGAA	1937
QY	1863	aaggaacagagaaaaaagaagaagagagaagaagagagaagcylgatagacygca	1922
Db	1938	AGAGAGGAACCTATTCAGAGAGAAACAGAGAGAAAAAGAAAGACGCCGTAGCAGAGAT	1997
QY	1923	gaggaagaatttgactaagatcttgcgcgcagtggttgaaaggagaagcagcaggaagga	1982
Db	1998	GAGCAGAAAGAGAAAGAAAGAGATCGTTAGAGACATATGAGAGATGAGCA-----AGCTA	2051
QY	1983	gagaagaattttagaagaatttagtcagycctcagacaagtoagggagacctgycgaatg	2042
Db	2052	TTGGCCACTGTGCTGTTAGTGGACAGAAACAGAGATATGACAGAGAGAGAAC-----GAAGAGG	2108
QY	2043	accaccactcgacaagaagccagtgctglatgtlaaagaanaagcaactgycgaagac	2102
Db	2109	TCCCAACTCGATTCGCGACACAGTGTGCTTATCGCAAGAAAGAAAGGGGCACTGGGCTAAAGAT	2168
QY	2103	tgccccaagaa-----ggagaacaagaagccgaagttcctagctttagaagaaga----t	2153
Db	2169	TGTCCCAAGAACACAGAGACACTCGGGACACAAAGCCCAACCTCCCTCCTGACCTTA	2228
QY	2154	aaagattagggagagcgggtgtcgaacccctcccgagcccgaggttaactltgaagtg	2213
Db	2229	GATGACTATGGGAGGTGACGGGTCAAGAGACCCCCCTGAAACCCAGGATTAACCTCAAGTC	2288
QY	2214	gaggggcaaccaggttgagttccctggttgataccggagcggagcaltcagtgctctaca	2273
Db	2289	GGGGGCAACCCGTCACCTTCCTGGTATGATATCTGGGGCCCAACACTCCGTCTGACCCA	2348
QY	2274	ccattaggaanaactaanaaanaaanaactcgtgtgatbgtgtgcacagggcacaagcgag	2333
Db	2349	AATCCTGGACCCCTTAAGTATGATGTCGTGGCTGGGTCCAAAGGGGCTACTGAGAGAAAGCG	2408
QY	2334	tatccatgagctacccgagaagaacccgttgacttgggaatggagcgggtaacccactcglt	2393
Db	2409	TATGCTGTGACACAGCGATCCCAAAAGTATCATCTAGCTACCGGTAAAGCTACCCACTCTTC	2468
QY	2394	ctgtgcatccctgagtgcccgagtaacccttctagtagtagagacttctgacaagaatgga	2453
Db	2469	CTCCATGTACCAAGACTGTCCCTATCTCTGTTTAGAAGAGATTTTCTACTAAACTAAAA	2528
QY	2454	gtccnaattctcttgacaagaagaaccagaagatgctcgtgaataaca--aaaccatc	2510
Db	2529	GCCCAATTCACATTTGAGGGATCAGAGAGCTGAGTTATGGGACCAATGGGGCACCCCTG	2588
QY	2511	actgtgttgccctccaatttagatgtagatctgacataatctccccaagtaaaagct	2570
Db	2589	CAGAGTTGACCCCTAAATATATAGAGATGAGCATTCGCTTACATGACGACCTCTCAAAAGACCA	2648
QY	2571	gatacaagatatacagltcc--tgttggagcaggttccccaagcctctggcggaaacgca	2627
Db	2649	GATGTTTCTCTAGGGTCCACATAGGTGTCGATTTTCCTAGGCGCTGGGGCGAAGCGGG	2708
QY	2628	ggagtggtttggcagaagcaagttcccccacaaggttatctcaactgaaagccagtgctaca	2687
Db	2709	GGCATGGGACATGACAGTGTGCCCAAGTCTCTTGATCATACCTGTGAAGCAACCTCTTACC	2768
QY	2688	ccaattacagctcagacagtaaccccttgagtagagagctctgagaagaatctggccgat	2747
Db	2769	CCGCTGTCCATAAACAATATACCCCATAGTCCAAAGAACGCAACTGTGGGATCAAGCCCAAC	2828
QY	2748	gttcaaaagatlaaccacaagggcatcagtgttctgtccaatccctctggaabactcc	2807
Db	2829	ATACAGAGACTGTTGGACCCAGGAATATAGTGTATCCCTGACAGTCCCTCCCTGGAAACAGCCC	2888

QY	2808	ctgttaccggtttgagagccttggaccaaattgtatctgaaccagtaacagactttgagag	2867
Db	2889	ctgtctaccctgctttagaanaaccaggactaatgtatattatagccctgtctccagagacttgagaa	2948
QY	2868	gtcaatlaaagagggtgacagacatacaaccacagtcctccgaacccttataaccctcttagc	2927
Db	2949	gtcaaacaaaggsggtggaaagacatccaccaccctgtagcccaaccttcaaacctcttgagc	3008
QY	2928	ggcctccgcgcctgaaggaacttgtacacagatattggaacttaaaagatgccttcttcgc	2987
Db	3009	ggcgtctccaccggtccaccacagtgctgactgtgcttgaattttaaagatgccttttctgc	3068
QY	2988	ctgagattacaaccaccactagaccacacacttctgccttcgaatgtgagagatccagttacg	3047
Db	3069	ctgagactccaccaccacacactagacactctcttgcgcttttgatgtgagagatccagatmg	3128
QY	3048	ggaagaaacgggacgtacactacacttggaccgcgaactgtccccaaaggtltaagaactcccagc	3107
Db	3129	ggaaattccaaagcaaatgtgactctggaccacacactccacaaaggttttcaaaaacagttccacc	3188
QY	3108	atctcttgacgaagcccttacaacagagacccgycgcacacttaagatcaacacactcaagtg	3167
Db	3189	ctgttttatatagcacttgcacagacactttagcagacttccgattccagaccaccacagcttg	3248
QY	3168	accctccacagtaagtlgatalgacactgtctcttgcgaggagccaccaaacaggaactgtcta	3227
Db	3249	atctcgtctacagatgacgtgagtgactttagctgacccgcacacttctgacactgacactgacaa	3308
QY	3228	gaagatgcaggaagcacactacgtcgtcgtgaattgtcttgacactgagtaacagacctctctaa	3287
Db	3309	caagcttactcgggccctctgttacaacaccttagggaaactcggtatcgggctctcgccacag	3368
QY	3288	aagcgcacaagatttgcagagagagagtaataacttggatagatttgcgaggcgaggag	3347
Db	3369	aaaccccaaatgttgccagaaacagtgcaagtatctgggctatcttcttcaaaagagggctcag	3428
QY	3348	cgatgtgtgcgagagagacggaagaaacttgaatccagataacgycgcccaacacagac	3407
Db	3429	agatgctgtgactagggccacaanaagagactgtatggggccagacttactccgaagaccctt	3488
QY	3408	aaacagattgagagagtttcttggagaaagcttgaatttgcagacttggatcccgagttt	3467
Db	3489	cgaaacttaaggagatttccttagggagacggcagggcttctgcctcgatcccttggtt	3548
QY	3468	ggcagctttagcagggcccaactctaccccgctaacccaagaagaaagggagattctctgggct	3527
Db	3549	gcacaaatggcagggcccttgccttgccttccacccaacgggacactgtgtttatgtggggc	3608
QY	3528	ccttagaccacgaaggaatttgaatctatcaaaaagggccctcttagcgacactgtctg	3587
Db	3609	ccaaactcaaaaagggccttctcaagaatcagaaagcttcttctaacttcccagacctg	3668
QY	3588	ggccctccctgaagtaactaaacccttlaccccttattgtagttagcgttaaggagtagcc	3647
Db	3669	gggtgtgcacgaatttgaacttagcccttcttgaaacttttgtgacagaaagcagggcttaccgc	3728
QY	3648	cgagagatttaaacccaaccctcagaacatctgagagagacccgttgcctactctgtcaag	3707
Db	3729	aaagctgtccttaagcacaataacttggagactctggcgtcgccgcccgttgactactgttcana	3788
QY	3708	aagcttatacctttagacagattgttgcgcgtatgtctgaaggtctatgcagactgttgcc	3767
Db	3789	aaagttagaccagatgacagctggggtggccctcttgccttgcgatgtgtagcacccatttgcc	3848
QY	3768	atactgttcaagagacgtctgacaaattgacttcttggagacgaatataactgtatagcccc	3827
Db	3849	gtactgtgcaaaagatgacgacacgctatccatgagcacacccactgtatcttctgccc	3908
QY	3828	catgacttggagaaacatcgtcttgcagccccacagacgaatgatatccaaacgcgcgaatg	3887
Db	3909	catgcagtaagagacactgaactcaanaaacaccccgacacgcttgcttccaagcccgagatg	3968
QY	3888	accacatatcaagcctgtcttctt---cacagagagggtacatttgccttccacacagcgt	3944

Db	3969	ACTCATATACAGGCGTTGGTTTGGACACGAGACCGGGCTCCATTGTCGACCGTGGTAGGCC	4028
Oy	3945	ctcaaccctgcacctctcttgcttgaagagactgatatgaaccagtgatctcattgattgcac	4004
Db	4029	CTGACACCGGCTACGGTGGCTCC - - ACAGCTGAGGAAGGGCTCCAAACACACATCGCTT	4085
Oy	4005	caactatgtatbgaagagacttgggtctcgcaagagacctatagacatacctgactgtga	4064
Db	4086	GATATCTGGGCCGACGCCACGSAACCCGACCCGACCTTAAGCGACACCGCTCCCAAC	4145
Oy	4065	gaagtgtctaaccttgatctcactgcaggaacagctatgtgtgtggaagttaagagatgtct	4124
Db	4146	GCCGACACCACTGGTGTACAGGGATGGAGAACAGTCTTTCAGAGAGGGGACAGCGTAAAGCG	4205
Oy	4125	ggggcgccagtggtgtgaacgggaccgcagcagctctgggccaagcctgtccggaaggaact	4184
Db	4206	GGAGCTCGGTGACCCACCGAGCCGAGGTAATCTGGGCTTAAGACCTTGCCACCGGAC	4255
Oy	4185	tcagcgcaaaaggctgtgactcactatgcccctacgcgaagcttttcggtctggccgaagga	4244
Db	4266	TCGCGTCAGGGGGGTGAATGATAGCACTACCAAGCCCTAAATATGGCAGAAAGTAAAG	4325
Oy	4245	tcataaactattatacgcgcacagtgatgtcctttggcactgcgcacogtacaogggcc	4304
Db	4326	AAGCTAATGTTTATCTGTATACCCCTTATGTCTTGTCTACGTGCCCATATCCATGGAGAA	4385
Oy	4305	atcctatacaaaagggtgtgtcttacctcagcaggaaggaataaagaacaagaagga	4364
Db	4386	ATATATCAGAAAGCGGTGGTGTCTCACATTCAGAAAGCCAAAGATCAAAAATTAAGACAG	4445
Oy	4365	attctaaagctattagaagccttacaatttgcacaaaaggctatgctattataacctgtctc	4424
Db	4446	ATCTTGGCCCTACTTAAAGCCCTTTCTGCCCCAAAGACTTAGATATCATCTATGTCTCA	4505
Oy	4425	ggacatagaagaagccaagatctcatatctagaagggaacagatgtgcagccggtctgc	4484
Db	4506	GGACATCAAAAGGACACACAGCGCCGAGGCTTAAGAGCAACCGGATGGCTGACCAACGGCC	4565
Oy	4485	aagaagcagcccaagctgtta-acctctgcctataatagaacgcaccaagcccca	4533
Db	4566	CGAAAGGAGGCATCACAGAGACTCCAGACACTCTTACCTCTCATAGAAATTCATCA	4625
Oy	4544	accacagacagtaaccccttaagaagcctgtgcgaaggaataaagaatagaccagttctc	4603
Db	4626	CCCTACACCTCAGACATTTTCATTAACAGTACTGATATTAAGACCTAACCAAGTTG	4685
Oy	4604	tgaagctccggag-----ggagactgtctaacctctatgggaaggaatactctgcc	4655
Db	4686	GGGGCATTTATGATTAACAAGAAGAGTATGGGCTTACCAAGAAACCTGTGATGCT	4745
Oy	4666	cacaagaagaggttagaatctgtccacaagatacatcgtctcaaccact-agaactaa	4714
Db	4746	GACCAGTTTACTTTTGATTATTAGACTTCTTCTTCACACTGACCTCACCTCAGCTTGCA	4805
Oy	4725	aacactcagcagtggtgtgaagaacatccctctcaatgtctcgaagctcacaagagtggc	4774
Db	4806	AAATGAAAGGCTCTCTTAGAGAGAACGCAAGTCCCTACTACATGCTGAACCGGATGGA	4865
Oy	4775	tgactcgtgtgtcaacaatgtgtgcctctgcagctcgtgttaatgtcaatcctctccagaat	4834
Db	4866	ACACTCAAAATATCACTGTGACCTCGCAAGCTTGTGCACAAGTCAACGCCACGCAAGCT	4925
Oy	4835	acc-----tcagaagaagaaactaaaggggaagccaccagcgagctcactcaggaagtgac	4889
Db	4926	GCCGTTAAACGGGAACTAGGGTCCGGGGCATCGGCCCGGCACTCATTTGGGAGATCAT	4988
Oy	4890	ttcactagagaaagcccggtctaataacggaacaaataltctatlttgttltgttagacac	4949
Db	4986	TTCAACGAGATTAAGCCCGGATTGTATGCTATAAATATCTCTAGTTTATGATATCA	5045
Oy	4950	tttcaagatgtgttagagcttatcctactaagaagaagacttcaacggtgtgtgttag	5005

Dh 5046 TTTTTCGGCTGGATAGAGCCCTTCCACCCAGAGAAACCCGCAAGGTCGTAAACCAAG 5105
Qy 5010 aaaaactcgaagaaattttccagaatttgaaatacctaagtaataaggtcgaacat 5069
Dh 5106 AACCTACTAGAGAGATCTTCCAGGTTCCGATGCTCAAGTATTTGGAACTGCAAAAT 5165
Qy 5070 gttccagcttctgctcccaagtaagtcaggaacttgccaagaatatttggaattg 5129
Dh 5166 GGGCCCTGCTTCTGCTCCAGAGTGAAGTCAAGAGTGGCCGATCTGTTGGGAAATGATTGG 5225
Qy 5130 aaactcatttgcatacagaaccccaagaactgaagaagtgtagaagataagaaacc 5189
Dh 5226 AAATTCATTGTTCATACAGACCCCAAACTCAGGCCAGGTAGAAAGATGCAATGAACCC 5285
Qy 5190 attaaagagaaccttaaccgaattgacacagagactgagtaagattgagtgctcc 5249
Dh 5286 ATCAAGAGACACTTAACTAAATTAACGCTTGCACATGCTCTAGAGACTGGGTGCTCTTA 5345
Qy 5250 ctgcaccttgcttctttagagtgaggaacacccctgagacagtttgagctgaacccat 5309
Dh 5346 CTCCCTTAGCCCTGACCGAGCCGCAACACGCGCGGCCCCCATGGCCTCAACCCCATAT 5405
Qy 5310 gaattcctcagcgagaccaccccgcttgagagaatgctccttgacataagtgctgat 5369
Dh 5406 GAGATCTTATATGGGACACCCCGCCCTGTGAACCTTCCCTGACCTGACATGAACAAGA 5465
Qy 5370 gtgcgtcttccagacctgttctcctagagtcgaagcgctcgaatgagtgagtgagcga 5429
Dh 5466 GTTACTAACAGCCCTCTCTCCAAAGCTCACTACAGGCTCTCTACTTACTTCCAGACACAA 5525
Qy 5430 ggttggaagcaactcccggaagaccttaactgaagaagagacttgaa-----gttccaat 5483
Dh 5526 GTCTGGAGACCTGTGGCGGACCCCTAACAAAGAACATGGAACCGCGGTGGTACTCTAC 5585
Qy 5484 cgtctccaagttggaagatcagctatgtitagaagcacaacgctgagagaactcgaagact 5543
Dh 5586 CCTTACGAGTGGGGGACACACTGTGGGTCGCGCCGACACACACTTAAGAACCTTGAACCT 5645
Qy 5544 cgttggaaggaacctatctcgtacttttgaaccacacaaagcgctgtgaaagtccaaga 5603
Dh 5646 CCTTGGAAGACCTTACACAGTCTCTGCTGACACACCCCGCCCTCAAAAGTGAAGCGC 5705
Qy 5604 atcccacactgatacagatccaggttaagccgc-----gcacact 5648
Dh 5706 ATTCGAGCTTGGATACAGCGCGCCACAGTGAAGGCTGCGAACCCCGGGGGTGAACATCC 5765
Qy 5649 cccgaactcgggttgagagacgaagaagactgagaatcccttaagcttcgcctcatcgc 5708
Dh 5766 TCTAGACTGACATGGCGGTTCAAGCTCTCAAAAACCCCTTAATAATTAAGTTAAACCCGC 5825
Qy 5709 ctggttctacttaacaataactcccaagccagtag-----taaacgcttataagaa 5764
Dh 5826 GAGGCGCCCTATCCCTTAATTTCTTGATGCTCAGAGGGGTGACTGCTTGCGCCG 5885
Qy 5765 gctcgaaccccatagagacttatacccttactcgtgctgattatagaccctgatacgggtg 5824
Dh 5886 GCTCAGTCTCATCAAGTCTATATATATACCTTGGAGAGTAAACCAATGGAATG---GGG 5942
Qy 5825 tcaactgtaataagacactcgaagtggtgctccttagaagagacacttggtgccttaac----- 5878
Dh 5943 AGACGATGAGGCAACTCTGCAACACCCCTGTGTGACCTGTGTGCTGTACTTACC 6002
Qy 5879 -----tgatttctgcctcgaatgattaaaccccgctgtataaagacacactccaa 5930
Dh 6003 CAGATTTATGTATGTATGACCCACCATGACCATTTATTTGGGGCTGGAATATCAATCC 6062
Qy 5931 cctagtc-----gtagtatagtgtctatactgctgcgaagagagaagaaga 5980
Dh 6063 CTTTTCCTTCCCGCGGGCCCTGTTGCTCAGGGGGGAGAGACCCAGCTGTTCCA 6122
Qy 5981 aatacttgaggtgtcctggaatcctt-----cgttagagaatlgag 6023
Dh 6123 GAGACTGGAGAAACCTTAACTCCTCCACCCCTGCTGCAACATGCTCGGAACAGAC 6182

Qy 6024 ctgcgtacactcacaagatgagactggaatatgscgatctctccagaacgggtlaa 6083
Dh 6183 TCMAAGCTAGACAGCAACTATATATCAAAATGAGGAGATTTATGTTGCCCGGGGCC 6242
Qy 6084 attctccttgtaacttcggcccgagggaatlaaagaatgtgaact----- 6131
Dh 6243 ACCGCCCGGAGATTCGAAGTCATGTGGGGTCCAGACTCTCTTACTGTCCTATTGGG 6302
Qy 6132 -----ataaagaataagagctgtctcccatcgaactgaattatc 6172
Dh 6303 GCTGTGAGACAACCGGTAGAGCTTACTGGAAGCCCTCTCATATGAGGATTTTCACAG 6352
Qy 6173 taagataagttcaactcgaagaagaacaggaataatlaaagaatgataatgata 6232
Dh 6363 TAAACMAAATCTCACCTGTGACAGGCTGTCCAGGTATGCAAAAGATATTAAGGTGCA 6422
Qy 6233 agctggggaatagttttatataatgagcgaggaagagaggttccacttaacattcgc 6292
Dh 6423 ACCCTTAGTTATTTCGGTTACAGACGCGGGAGACGGGTTACTTCTGACACAGGAC 6482
Qy 6293 ttaggataagagcgaggagaaacccctgtgcaatgggaacccgataaagtactgctg 6352
Dh 6483 ATTACTGGGCTTACGTTTGTATGTCGCGGACAAGATTCAGGGCTTACATTTGGATTC 6542
Qy 6353 aacaggagcccgcccgccctgagccac-----gataactgacggtgccccaatcaact 6409
Dh 6543 GACTCAGATTCACAAATCTAGAGCCCGCGCTCCAAATGAGGCCAAACCCGTTCTGGCAG 6602
Qy 6410 cgttgagcctgacataaacaagcgcctagcaacagtaacaactgattatctacaa 6469
Dh 6603 ACCAAAGCCGACCTCCAGGCCCAACCTGTAAAGTGGCTTACGTACCAACACACCA 6662
Qy 6470 acaagcctagaactcc-----agggttctcgttgaagaagaagaagacttca 6523
Dh 6663 GTGGAGCTCTCTCTCCCTACCAACCTTCACCGCGGGAGAGGAAATATGAGCTGTAA 6722
Qy 6524 gtccatcagaagagcttccaagcactcaactccaccccgctgagccacttctct 6583
Dh 6723 ACTTAGTAGAGAGAGGCTTACCAGCCCTCAACCTCACCACTCTGACAAAAACCAAGAT 6782
Qy 6584 gttggttcttctactcctcagggcctcttattatgagggagatggtcctaaagaagaat 6643
Dh 6783 GCTGGTTGTCTGTAGTGGGACCCCTCTACTAGCAAGAGGGTTCCTGCGGTCAAGCT 6642
Qy 6644 tcaatgtgaccaaagagcatagaataatgatacgtcgtgaggtggtccgaataagcttaacc 6703
Dh 6843 ACTCCAAACATPACCTCTGCTCCAGCCCAAGCTCCGCTGCCCTCCAAACACAAAGTTGACC 6902
Qy 6704 tcaactgaagttccgggaagggagacatgcatagagaagaagcttcccatcccaacaacac 6763
Dh 6903 TGTCCGAAGTAGACCGGACAGGACTGTGCTATAGAGACAGTTCCTCCAAACACATCAAGGCC 6962
Qy 6764 ttgctatagtagtctggttattaggaagcctcagaataatcagtaattagtaactggtt 6823
Dh 6963 TATGTATATACACCCAGACAGCAAGCATGAGGCTC-----CTATATATATGTTGCCCTTA 7016
Qy 6824 ataacaggtggtgagatgataactggtttaaaccctggttccactcaagtctta 6883
Dh 7017 CAGGTACCATGTGGGCTGTGATGACGGGCTTACTCATGACATCTCCACCATCATCTCA 7076
Qy 6884 accaatcaagaatttctgtaatgtaacaaatgctcccgagtgtaactcaactccg 6943
Dh 7077 ACCTTACCACTGATATTTGTTGTTCTGTGGAACCTTGSCAAGAGTCCACTTATCATTTCC 7136
Qy 6944 aggaagatggtccttgatgaataactatcgtatataacggaacaaagaagaacggtat 7003
Dh 7137 CCACTATATTTAACGGCTGTTTG---AGAGATTCACACGACACAAAGAAAGAACGGGTGT 7193
Qy 7004 ccttaaccctagctgaatgctcgaatlaggaagcgctgtggtgtagaagaagaagaag 7063
Dh 7194 CGTTAACCTGGCCCTATTTATTTGGGTGAGCTAACCATGGGGGAAATTTGCCGTGGAATAG 7253

QY 7064 ctgacctgatacagagcacagcagctagagaagagcttgtagctacatgcggcca 7123
 Db 7254 GAACAGGAGACTGCTGCTTAATGGCCACTCAGCA---ATTCCAGCAGCTCCAAAGCCGCG 7310
 QY 7124 tgacagaagaatctccagacttaaaagagctgttagcaactagaagaagctccagactt 7183
 Db 7311 TACAGATATCTCAGGAGGTTGAATAATCAATCTTAACCTAGAAAAGTCTCACATT 7370
 QY 7184 cttgtctgaagtgttctacagaaccggaaggagatagatctgctgttcttaagaaga 7243
 Db 7371 CCTGCTCTGAAGTTGTCTTACACAAATGGAAGGGGCGCTTACACTTGTATTCTTAAGAAG 7430
 QY 7244 gtgtgtatgtgacgacttaaaagaagaatgtgtcttcagttagatcactaagaagcca 7303
 Db 7431 GAGGCGTGTCTGCTCTTAATAAAGAAATGTTGCTCTTATGGGACACAGAGACTAG 7490
 QY 7304 tcagagactcatalgaacaagcttagaanaaaagtlagaagcgctcgaaagggaagaagag 7363
 Db 7491 TGAGAGACAGCATGGCCAAATGAGAGAGAGGCTTAATCAGAGACAGAAACTGTTGAGT 7550
 QY 7364 ctgacacagaggtgtttagaagagtgltcaacaggtctcttgatgagaccocctgctt 7423
 Db 7551 CAACCTCAAGAGATGTTTGAAGGACTGTTAACAGATCCCTTGGTTTACCACTTGAATAT 7610
 QY 7424 ctgctctgacgggcccctagtagtctgctgcttacttaacagtlgggacctgtttaa 7483
 Db 7611 CTACCATTAATGGAGACCCCTCATGTACTCTTATGATTTTGTCTTGGAGCCCTGATTC 7670
 QY 7484 ttaataagttgttgccttctgttagaagaacagagtgagtgcaagtcagatcagtgactta 7543
 Db 7671 TTATCATTAATGATCAATTTGTTAAAGACAGATATCAGTGTGTCAGGCTTATGTTTGA 7730
 QY 7544 ggcacacagttacca 7556
 Db 7731 CTCACAAATATCA 7743

RESULT 12
 US-08-110-300A-9/c
 ; Sequence 9, Application US/08110300A
 ; Patent No. 5643756
 ; GENERAL INFORMATION:
 ; APPLICANT: Pinter, Abraham
 ; APPLICANT: Kayman, Samuel
 ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
 ; STREET: 45 Rockefeller Pl.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/110,300A
 ; FILING DATE: 20-Aug-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jacobs, Seth H.
 ; REGISTRATION NUMBER: 32,140
 ; REFERENCE/DOCKET NUMBER: 11698A50
 ; TELEPHONE: 212-757-2200
 ; TELEFAX: 212-586-1461
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 10367 base pairs
 ; TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOHEATICAL: NO
 ANTI-SENSE: NO
 US-08-110-300A-9

Query Match 19.4%; Score 1596; DB 1; Length 10367;
 Best Local Similarity 58.4%; Pred. No. 0;
 Matches 3161; Conservative 0; Mismatches 2135; Indels 114; Gaps 17;

QY 467 tccgagcgagctccagactcttgcctgctgtgaa--gacggagcgggtcgctgtgt 524
 Db 7098 TCCGCCGCCCTGTGAGTTGCTTGGTTGGTAACCGAAGCCGCCGCCGCTCTTGT 7039
 QY 525 ctgagatctgtgttctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 584
 Db 7038 CTGCTGACGACATCGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6981
 QY 585 aataggagacagagtagtactaccccccttagttagctcagacattgagctgaagt 644
 Db 6980 AACATGGGCGACGCTGTATACCAACCCCTTAAGTTGACTTAAAGCACTGAGAGATGTC 6921
 QY 645 agatcagaggtcatabaattgtcagtlcaggtltaagaaggacttggcagacttctgt 704
 Db 6920 GAACGACAGGCCAACACCTGCTGTAGAGGTTGAAAAAGCCCTGGTTTACATTTCTGC 6861
 QY 705 gctctgaatggcagacattgagtgatgagccatcagagaggagacttaattctgaa 764
 Db 6860 TGTGCAAGATGGCCAACTTCAACGTGCGATGGCCAGACAGGACACTTTAAACCGAC 6801
 QY 765 attatccctgctgttgaagcaatcatltaagaactgagccgggtcctcactatcag 824
 Db 6800 ATTATTACAGAGTTAAGATCAAGSTCTTACACTGGCCCAATGACATCCGAGATCAG 6741
 QY 825 gagccctatctccttaagtgagcaaatltagcagaagatcctccgagtggttaacca 884
 Db 6740 GTCCCTCATCATGTGACCTGGGAAGCTATACAGTAGACCCCTCTGCTGAGACCC 6661
 QY 885 tggctaaataaa-----cgaagaagcaggtgccccaatccgtg 925
 Db 6680 TTCTGTGACACCTTAACCTTCCCTCTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 6621
 QY 926 tcttgagagaaacaaacactcgccggaagaaagtgagcccttctcgtatcacc 985
 Db 6620 CCCCCACTCTCGACCCCGCCCGCCAGCTCTCTATTCGCGCTTCTACTTCTCTTTAAG 6561
 QY 986 ccgagatcgagagcgcgcgacttgcggaaccccaacccgttccccaacccctatc 1045
 Db 6560 ACCAATCTTAGGCTCTCAAGTCTTCTCTATAGCGAGGACCACTATGATCTACTCAG 6501
 QY 1046 cagcagaggtgtgttgagagacacttgcctcctcgtgagtcctcggtgtgtgagagac 1105
 Db 6500 GAGGACCCCTCCGCTTACCGGACCCAGGACCCCTCTCTGAGGAGGAGCGCATAGC 6441
 QY 1106 -----tgcgcggagactcgagccgagagagggcgccaccccggaaggagcagagaga 1158
 Db 6440 GGAGAGTGGCCCTTACAAAGAGAGCCCTTGACCTTCCCAATGCTATCCCGCTGCG 6381
 QY 1159 tcgagatattacgcgtcgccacttgcgcctcccatgcagagggggccaattgcagccc 1217
 Db 6380 GGAAGAAAGAACCCCGCTGCGGATCTACTACTCTCAAGCGCTTCCCTTCCCTTCCCT 6321
 QY 1218 -----ctcagataggcccttctcgtcagatcctataatggaaa 1262
 Db 6320 GGAGGAATGACAGATATACATAGCGCATTTCTCTCTGACCTTATTAACGGAAA 6261
 QY 1263 actaacatccccccttctcggagatcccccaagccctcaggggtgtgtgagtcctt 1322
 Db 6260 AATAACACCCCTCTTCTTCCAGAGCCAGCTTAATTACAGCTTGTATGATGAGTCGCT 6201
 QY 1323 atgtctcaccagcctacttgagatgtgtlcaacagctgtcgtcagagactcttca 1382

Db 1833 AGCAGATCGACATGGGTGTTCAACGCTCCCAAAATCCCTAAGATTAAGATTGACCG 1774
QY 5708 ccttgcttactactaacaactaactccagcagtagtaacgctatagacagct 5767
Db 1773 CGGAGCTCTTAATCCCTTAATTCCTCTGCTCTCAAGGGGCGCAATCCGACGCA 1714
QY 5768 cgaaccccca 5777
Db 1713 CCGGCTCCA 1704

RESULT 13
US-08-886-642-9/c
Sequence 9, Application US/08886642
Patent No. 5952474
GENERAL INFORMATION:
APPLICANT: Pinter, Abraham
APPLICANT: Kayman, Samuel
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,642
FILING DATE: 01-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/110,300
FILING DATE: 20-AUG-1993
APPLICATION NUMBER: 07/938,100
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 07763/010002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212/765-5070
TELEFAX: 212/258-2291
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10367 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-886-642-9

Query Match 19.4%; Score 1596; DB 2; Length 10367;
Best Local Similarity 58.4%; Pred. No. 0;
Matches 3161; Conservative 0; Mismatches 2135; Indels 114; Gaps 17;

QY 467 tcggggccgctcagactcttctgctcgtctgctgaa-gacggagagcgctgctgct 524
Db 7098 TCCGCCCCCGCTGAGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT 7039
QY 525 ctgacatctgtgttctctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 584
Db 7038 CTGCTGAGCATGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 6981
QY 585 aatatggagacagactgactacccccctagtgtgactctcgacatctgagctgagtt 644
Db 6980 AACATGGGCGAGGCTTTTACACCCCTTAAGTTTGAAGTTTGAAGTTTGAAGTTT 6921

QY 645 agatcagggctcataaattgtcagttcagtttaagaaggaacttgcagacttctgt 704
Db 6920 GAACGGACAGCCACACACTGTCGTAGAGTTTGAAGAAAGCGGCTTACATTTCTGC 6861
QY 705 gctctgaatgagcaacatctgattgtgattgacatcagagggagcccttaattcgt 764
Db 6860 TCTGAGAAATGGGCAACCTTCAACGTGGATGGCCAGAGAGGCGCACTTTTAACCA 6801
QY 765 attatcctgtctgttaagcaatcttctcagactgagccggcctctatctgatt 824
Db 6800 ATTATTACACAGGTTAAGATCAAGGCTTCTCAACCTGGCCACATGAGCATTCGGAT 6741
QY 825 gaacctatactcttgcagtgagcaagattggcagaaatctccgcagatggttaacca 884
Db 6740 GTCCCTTACATCGTGAAGCTGGAGGATTAAGCATGAGACCCCTCTCCCTGGGTCA 6681
QY 885 tggctaaataa-----ccaagaagccaggtccccaatcctgtgc 925
Db 6680 TTGCTGACCTTAACCTTCCCTCTCTCTTCCCTTCAAGCCCTCTCTCCACCTGAA 6621
QY 926 tcttggagagaaacaaacactcgcgcgaagaaagtcgagcctctctctgtatctac 985
Db 6620 CCCCACCTTCGACCCCGCCCGAGCTCCCTCTATCCGCTCTCACTTCTCTTAAAC 6561
QY 986 ccgagatcgagagccgcgacttgcgcggaacccaactgttccccaacccctatc 1045
Db 6560 ACCAAACCTTAGGCTCAAGTCTTCTGATGAGGAGGACCACTCATGATCTACTACAG 6501
QY 1046 cagcagaggtgtctgtgagggagccctcctcctctgagctcgtgtgtgagggacc 1105
Db 6500 GAGGACCTCTCGCTTACCGGGAGCCAGCCACCCCTCTCTGACGGGAACGGCGATAGC 6441
QY 1106 -----tgcgcgagactcgcgcgagagcgccaccccgagcgagacagagaga 1158
Db 6440 GGAGAGTGGGCCCCCTACAGAAAGGAGCCCTGACCTTCCCAATGATTCCTCCGCG 6381
QY 1159 tgcgataatcgcgcgcgcacatagccctcccaagcagggagccaatctgagccc 1217
Db 6380 GGAAGAAAGAACCCCGCGGCGGANTTACTACTCTCTCAAGGCTTCCCTTCCGCTGC 6321
QY 1218 -----ctccagatctggcccttctctctctgagatctctataattgagaa 1262
Db 6320 GGAGGAAATGAGACATCAATTAATGAGCAATTTCTCTCTGATACCTTAACTGAGAA 6261
QY 1263 actaacatcccccttctcgcgagagatcccaacgcctcaaggggtgtgtgagctcct 1322
Db 6260 AATAACAACCCCTCTTCTCTCGAGAGCCACCTAAATTTGACAGCTTTGATCGAGTCCG 6201
QY 1323 atgtctctcaacgactacttggatgtgattgtcaacagctgcgcgaagacacttca 1382
Db 6200 CTCTTACTCATCAAGCCCTTGGATGACATGCAACAGCTTATGAGGACCTCTCTACG 6141
QY 1383 accgagagagagagagatctgttgaagctclagaanaaaatgtctcgtggcgaggg 1442
Db 6140 GGAGAGAAACACAGAGTCTCTCTAGAGGCCGGAAGAGGGGTTCAAGGGGAGAGAGGA 6081
QY 1443 cgaaccagcagttgcaaatgagatlgacatlggaattcccttgaactgcgcggattgg 1502
Db 6080 CGCCCAACTCAGCTGCCCAATGACATTAATGATCTTTCCCTTGGAACGTCGACGTGG 6021
QY 1503 gactacaacagcgcgtgaaggtgagagagcttgaanaatctctgcagagcctctgtgg 1562
Db 6020 GACTCAACACCCCAACGAGGTAAGGAGCAACCTAGTCAATTCGCGCATTTCTCTAGCG 5961
QY 1563 ggtctcgggggagcctcaagcagcgcacatattgtctlaagtgtaagagagtgatgag 1622
Db 5960 GGTCTTCCAAACGCGGGAGAGAGCCCAACATTTGGGCAAGGATTAACCAAG 5901
QY 1622 ggaacgaagaaactccctcggtattctctlgagagctcatgagagccttaagcggttct 1682
Db 5900 GGACTTAATGAGTCTCTCTCAGCCCTTTTAAAGAGACTCAAGAGGCGCTTATCGAGATAC 5841

QY 1683 acccttltgactccactcagagcccaagaagccctcagctgagccctcattgg 1742
Db 5840 ACTCTTTATGACCTTGAGGACCCAGGGCAAGAAACCAATGTGCCATGTCTATCTGCG 5781
QY 1743 cagtgcgtcttgatatacagaagaactcagagcagcggagaggttaccagaggtcag 1802
Db 5780 CAGTCCGCCCGGATATGCGGCGAAGATTAGACGGCTTAGAAGATTGGAAGATAGAGACC 5721
QY 1803 ttaagtcctagtgagagagcagaaggtgtattcagaagaaggagacagaagagag 1862
Db 5720 TTAGAGAGACTTAGTGAGGAGAGCTGAAGATCTTTATATAACAGAAACCCCGGAGAA 5661
QY 1863 aagaagaacagaaaaaagaagaagaagaagaagaagaagaagaagaagaagaaga 1922
Db 5660 AGAGAGAGACTTTATTAGAGAGAAACAGAGAGAAAGAACCCCTAGGCGCAGAGAT 5601
QY 1923 gagaagaatttgactaagatcttgccaggtgtgttgaaggaaagcagcagagaga 1982
Db 5600 GTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGATAGACAGGAGAGAGAGAGAGAGAG 5551
QY 1983 gagaagattttagaanaaattagtlcagccctagacagtcagga--acctggcaatag 2041
Db 5550 GTTGTGCTACTGCTGCTTACCGGGCAGAGACAGATAGACAGGAGAGAGAGAGAGAGAG 5491
QY 2042 gaccacactcagacaagagcaggtgtgtatttgaagaagaagaagaagaagaagaaga 2101
Db 5490 GCCCACTGACACCAAG 5431
QY 2102 ctgcccagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2152
Db 5430 TTGCCCCAG 5371
QY 2153 taagaatttagagagagagaggtgtcagacccctccagagccagaggaatttgaagt 2212
Db 5370 AGAGAGATTAG 5311
QY 2213 ggaaggcacaagcttgagcttggttgaacagagcagagcagcagcagcagcagcagc 2272
Db 5310 CGGGGGGCAACCGGTACTCTTCTAGTGAATCTGGGGCCCAACATCTGCTGAGACCA 5251
QY 2273 accattaggaagaacaaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2332
Db 5250 AAATCTCGAGACCCCTTAAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5191
QY 2333 gtaacatagagacacagagagacgttgacttgagagagagagagagagagagagagag 2392
Db 5190 GTATCGCTGAGACCAAGAGATGCGAGTGCATCAGGACCGGCTGAAGTGAAGTGAAGTGA 5131
QY 2393 tctgtatcccttgagtgccagtlacccttctaggttagagacttaccagaagatgg 2452
Db 5130 CTTCCATGTACAGATTTGGCCCTATCTCTGCTAGAGAAAGATTTGCTACTTAACCTAA 5071
QY 2453 agctcaaatcttcttga---acaagaagaacagaagtgctgtgataaacaacacat 2509
Db 5070 AGCCCAAAATTCATTTGAGGAGATCAGAGCTCAGGTTGGAGCAATGGAGACACCCCT 5011
QY 2510 cactgtgtgacccctcccaatttagatgataatagctatacttccccaagtaaacg 2569
Db 5010 GCAAGTGTGACCCCTTAATATAGAAATGATGATGCGCTACATGAGACTCTCAAAAGGGCC 4951
QY 2570 tgaataaatalacagtc---tggttgagagagcttccccaagcttgagagaagaagc 2626
Db 4950 AGATGTGCTCTAGGGTCCACATGCTCTCTATTTTCCCAAGGCTTGGGCGAAGACGG 4891
QY 2627 agggatgggttgagcaagaagctccccaaggtatccaactgaagcaggtgtac 2686
Db 4890 GGGGATGGGGCTTCCGCAAGCTCTCTGATCATACCTTGGAAGGCAACCTCTAC 4831
QY 2687 accaatatcagcag 2746
Db 4830 CCGGTGTGCTAATAAACAATACCCATGTCTACAGAAAGCAGAGCTGGGATCAAGCCCA 4771
QY 2747 tgttaagaatataccacagagagcagctagttcctgtcaatcccttgaaatacc 2806
Db 4770 CATACAGAGACTGCTGATAGGAAATCTGATACCTGCACTCCCTGGAAACCGCC 4711
QY 2807 cctgtaccaggtttagaagaagccttggaaccaatgatactacagagatcagagatgagaga 2866
Db 4710 CTTGATACCCGTTTAAGAAACCGGGAGCTATATGATTTATAGCCGTGTCCAGATCTGAGAA 4651
QY 2867 gttcaataaagggttcagagacataccaacaggtccccaaccttaacacctttag 2926
Db 4650 AGTCAACAGCGGGGTGAAGACATCCACCCACCGGCCAACCCCTTACACCTCTTGAG 4591
QY 2927 cgcctccgccttgaaacag 2986
Db 4590 CGGGCTCCACCGTCCACAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 4531
QY 2987 cctgagattacacccac 3046
Db 4530 CTTGAGACTCCACCCACAGTACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4471
QY 3047 ggaagaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3106
Db 4470 GGGAACTTCAAGCAATTTAACCCTGAGACAGACTCCGAGGGTTCAAAACAGTCCAC 4411
QY 3107 catctttagaagcctcacaag 3166
Db 4410 CTTGTTGATGAAGCCCTGACAGAGGACTCGCAGACTTCGCGATCCAGACCCAGACT 4351
QY 3167 gacccctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3226
Db 4350 GATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4291
QY 3227 agaagtaacagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3286
Db 4290 ACAAGTACGGCGGCGCTGTTTCAAAACCTTAGAGGACTTCGATATCGGGCTGCGCA 4231
QY 3287 gaagcccaagatttgcagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3346
Db 4230 GAAACCCCAAAATTTCCGAAACACAGTATCTGAGGATGATGATGATGATGATGATGATG 4171
QY 3347 ggcagcttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3406
Db 4170 GAGATGCTGATGAGGCGCAAGAAAGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGG 4111
QY 3407 caaacaagtgaagagatttcttgagcagcagcagcagcagcagcagcagcagcagcagc 3466
Db 4110 TCGACAACTAAGGGAGTCTTAGAGGAGCGAGGCTTCTGTGCTCTGATCCTGGGTT 4051
QY 3467 tgcagccttagcagcccaacttaccgctaaccaagaagaagaagaagaagaagaagaagaaga 3526
Db 4050 TCGAGAAATGCGACGCCCTTGTACCTCTCACCAAAACGGGAGCTGTTTGAAGTGGG 3991
QY 3527 tccctgaacacagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3586
Db 3990 CCGAGACCAAGAAAGGCTTACCAAGAGATCAACAGGCTCTTAACTCCCTGCGCT 3931
QY 3587 ggcctcccttagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3646
Db 3930 GGGATTCGACACTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 3871
QY 3647 ccgagagaggtttaaaccac 3706
Db 3870 CAAAGGTGCTTCAACGCAAAACCTGAGGCTTGGCGTGGCGGCTGACTGTCACAA 3811
QY 3707 gaagcttgatccttgtagcaggttggttgccgctatgtctgaaggtcctacgagctgtgagc 3766
Db 3810 AAAGCTTGAAGCCAGTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 3751
QY 3767 catactgtgacaagagcgttgacaatgtgacttgagagagagagagagagagagagagagag 3826
Db 3750 CGTTCTGACCAAGAGAGCGTGGCAAGCTCAACATGGGAGACCACTGATATTTGAGCC 3691
QY 3827 ccatgtatgagaaacatcgttcgagagcccaagacagatgtgatacagagcccgat 3886
Db 3691 CCGTTCTGACCAAGAGAGCGTGGCAAGCTCAACATGGGAGACCACTGATATTTGAGCC 3691

Db 3690 CCATGAGTAGAGGCACTAGTTAAGCAACCCCTGATCGTCTCCACGCCCGAAT 3631
QY 3887 gaccagatcaaaagccgtctct---cacagagagggtcaacttcgtcccaacagcgc 3943
Db 3630 GACCACCTACCAAGGCTCTGCTTCTGACACGACGACGAGTCCAGTTGGACCAATAGTGGC 3571
QY 3944 tctcaaccctgcacactctctcgtcgaagaagactgataagcaagtgactcaatgttgcca 4003
Db 3570 CCTAAACCCAGCTACGCTGCTCCCTCTA---CCTGAGGAGGAGGCTGCAACATCAGTCCCT 3514
QY 4004 tcaactttgtagaagagactggtgcgaagagacttaacagactacacgtctgactg 4063
Db 3513 TGACATCTTGCTGAACCCCAAGCACTAAGATTTACGAGACAGCTCTCCACGA 3454
QY 4064 agaagtgtaaccctgctcgtcgaagcaagctatgtgtggaagtaagaagatgac 4123
Db 3453 CGCTGACCACACCTGATACACAGATGGGAGAGCTTCTGCAAGAGGAGGACGCAAGGC 3394
QY 4124 tgggagggcaagtgtgagagggagccgaagctctgagcgaagcctgcccgaagaa 4183
Db 3393 CGGAGCAGCACTAAGCCACCGAGCGAGTATGCTGGGCAAAACCACTGCGAGCGGAGC 3334
QY 4184 ttgagcgcaaaagcttagctcatgcccacgcaagcttgcggtgagccgaagaa 4243
Db 3333 ATGCGCCAAAGAGCTGAGTTGATAGCCTACCCAGCCTTAAATTTGGCAGAGGTAA 3274
QY 4244 atccataacattatagcgaagcaagtgatgtccttgcagctgacagctgacagggc 4303
Db 3273 GAGCTGAAATGTTTACACCGATAGCGGTTATGCTTTGCCACTGCCATVATTCAGGAGA 3214
QY 4304 catctatacaaaaggggtgtgtctactcgaaggaaggaataaagaacaaagga 4363
Db 3213 AATATATGAAAGCGCGGGTGTCTCATCATCAAGAGAAAGAAATCAAAATTAAGACGA 3154
QY 4364 aattcaagcctattagaagccttaacatttgcacaaagagcttaactatatactgccc 4423
Db 3153 GATCTTGGCCCTACTGAAAGGCTCTCTTCTGCCAAAGACTTTGCAATTTCAATTTGCC 3094
QY 4424 tggacataagaagcaaaagatctcaatactagaaggaacagatgtgcggtgtgc 4483
Db 3093 GGGACATCAGAAAGGAAACCGCGCGAGGCAAGGGGCAAGAGTGGCGCAAGCGGC 3034
QY 4484 caagagcgagccagcagctgttaacctctcctataatagaagagcccaagccca 4543
Db 3033 CCGAAGAGTACGCTAGAGAACTCCAGAGACTTCCACACTTCTGATAGAAATTCGCGC 2974
QY 4544 accc-agaagcagctacacccctagaagactggaagagataaaagaatagccagttct 4602
Db 2973 CCCCTATACCTATGAAATTTTCATCTATACGGTGTGACTGACATTAAGATCTGACTAACT 2914
QY 4603 ctgagactccggaagggacc-----tgtatatacctcataatggaagaaatctgccc 4654
Db 2913 AGGGGCACTTATGAGAGTGAAGAGAGAGTGGGTTTATCAGGAGAAAGCCTGTAATGCC 2854
QY 4655 ccacaaagaaggggttagaatatgtccaaacagatacatcgtctaaacccctagaagaa 4714
Db 2853 TGATCATCTACCTTTGAACATTTAGATTTCTTCATCAATTTGACCCACCTCAGTTCTC 2794
QY 4715 acaactgacagactgtgtcagaaca-----tccctatacctatgtctgaggtcacaag 4768
Db 2793 AAAAACAAGGCTCTTCTGAGAAAGAACTAGTCTTATTCATGCTGGAACCGGGAGTCG 2734
QY 4769 agtgggtgactcggtgtcaaatatgtgtccctgagcagctggttaagtcaatccttc 4828
Db 2733 AACGCTCAAGACATCTAGAGACTTTGCCAAGCCTGTGACAGGTCAATGCCAGCAAGTC 2674
QY 4829 cagaatacctccagaaagaaactaaagggaaagccacccagcgctcactggaagtgga 4888
Db 2673 TGGCGTCAAAAGAGGAGTAAAGTTGAGGGACCGACCGGACCCACTGGGAATATGA 2614
QY 4889 cttaactgagtgtaaacggcgctaaataagaagaacaaatctatgtgttttgtaaacac 4948
Db 2613 TTTCACCTGAGTAAACCTGGCTGTATGGGTATTAATATCTTTTAGTTTTCATAGACAC 2554

QY 4949 ctcttcagagatggttagaggtctatccctaaagaagaagactcaacgggtgtgctaa 5008
Db 2553 TTTCCTGTGATGGTAGAAGCTTTTCCCAACCAAGAAAGAACTGCCAAAGTTGTAAACCA 2494
QY 5009 gaaatactcgtggaagaatttctccaagatttgaatlaacaaagtaagtaaggtcagaac 5068
Db 2493 GAAAGCTACTAGAAAGAAATTTTCCAGATTCGGATTCGCACAGATTTGGGAACCGACA 2434
QY 5069 tggctcagcttcgttgcctcaggttaagtgagactgtgcgaagataattgggtattgtg 5128
Db 2433 TGGGCTGCTGCTTCTGCTCCAAAGTAACTGACAGTAAAGCGGATTTATTTGGGCTGATG 2374
QY 5129 gaaactgacttctgac 5188
Db 2373 GAAACTACTAGTGTGCTTAAAGACCCAGAGTTTACGCTCAGTAAAGAAATTAAGAGAC 2314
QY 5189 catlaagaagacccctaccacaaatlgacacacagagactggaatlaatgattgtgctct 5248
Db 2313 AATCAAGAGAGACTTTAACTAAATTTAGCGCTTGCACCTGCTAGGAGCTGGGTGCTCT 2254
QY 5249 cctgccttctgcttctttaggttagaggaagacccctgacagatttgggtcgaacccca 5308
Db 2253 GCTTCCCTTGGCTTATGAGCCCGCAACGCGCGGCGCCCGACGCTCAGTCAACCCATA 2194
QY 5309 tgaattgctctacgagggagaccccccgttgcagagaattgtccttgcacatagtgtcga 5368
Db 2193 TGAATCTTATATGGGGACACCCCGCCCTTGAATCTTCCGTGATCTGACATGGCAAA 2134
QY 5369 tgtctgcttcccaagccttctctcagctcgaagcgcctcagttggttgaagagag 5428
Db 2133 GGTACTCATTAACCCCTCTCTCCAGCTCATTTTACAGGACCTTACTGCTGCGACGACA 2074
QY 5429 aggttgaagcagctcgcggaagcctactcag-----gaggaagcttgaattccaca 5482
Db 2073 AGCTGGAGACCGTTGGCGGCACTTACCAAGAGCAACTAGACCGCGGTATGACTCTCA 2014
QY 5483 tgccttcaagcttgaagatcagctatgtatgttgaagccacacgctgcaggaacactgagac 5542
Db 2013 CCGTTCGGGGTGGGTGACACAGTGGGTCGCCAGACACCAAACTAAATATTAAGAAC 1954
QY 5543 tgggttgaagagcctatctctgacttctgacacacacacacacacacacacacacacac 5602
Db 1953 CCGCTGGAAGAGACCCCTTACCGCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1894
QY 5603 aatcccaactggaatccatccatccacgttaagccggc-----gacacac 5647
Db 1893 CATTTGACGCGTGGATCCACGCTGCGCAAGTAAAGGCTGCGGACACCGATGAGCCAC 1834
QY 5648 tcccgaactcgggttgaagacccgaagaagactggaatcccttaagacttgcctccatcg 5707
Db 1833 AGCAGAAATCGACATGGGCTGTTCACACCTCCCAAAATCCCTTAAAGTAAAGATTGACCG 1774
QY 5708 cctgttcccttactctaaataactcccaagcagctagtaaacacacttaagacagct 5767
Db 1773 CGGAGCTCTTAATCCCTTATTTCTTCTCTCTCTCAAAAGGCGCAGATCCGACGA 1714
QY 5768 cgaaccccca 5777
Db 1713 CCGGCTCACA 1704

RESULT 14
PCT-US93-08041-9/c
Sequence 9, Application PC/TUS9308041
GENERAL INFORMATION:
APPLICANT: The Public Health Research Institute of the City of
APPLICANT: New York, Inc.
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Davis Hoxie Faithfull and Hapgood
STREET: 45 Rockefeller Pl.

QY	1963	gagagagattltagaanaattaggtacagccctagacagtcaggga-accttggacaatag	2047
Db	5550	GTTCGTGGCTACTCTCGTTAGCGGGGAGACACAGGATTAACAGGGGAGACCGAAGAG	5491
QY	2042	gaccccactgcacaagagaccagtcgtgacatctttaaagaanaagcacatctggacaaga	2101
Db	5450	GCCCCAACTGCACACACCAAGTGTGCTTACTCCAAAGAAAGGACATTGGGCTAGAGA	5431
QY	2102	ctgcgcccaagaaggaaacaagagaccgaagtcctagctctagaagaag-----a	2152
Db	5430	TTGCCCAAGAAAGCCAGAGGAGACCCGGGAGACACAGACCCAGCCTCCCTCAGCCT	5371
QY	2153	taagaataggggagagacggygtctcgagacccctcccgagcccaaggttaacttgaagt	2212
Db	5370	AGAGGATTAGGAGAGCTAGGGGTACGGAGCCCCCCCTCGAACCCAGGATTAACCTCAGAGT	5311
QY	2213	ggaaggagcaaccagttgafttcctggttttagccgaagcgagacattgaigtctctaca	2272
Db	5310	CGGGGGGCAACCCCTCACCCTTCCTAGGGATCTGGGGGCCCAACCTCCGTGCTACCCA	5251
QY	2273	accattaggaanaactaaagaanaaaatcctctgtgctgcatgggtgcacaagggacaagca	2332
Db	5250	AAATCTGGAGACCCCTTAGTGACAAAGTCTGCTGGGTCCAAAGGGGCTACTGAGAGGAGACG	5191
QY	2333	gtatccatgtgactaccggaaagacggttgacttggagtgggacggytaaccactgct	2392
Db	5190	GTATCGCTGGACCCAGGATGCGGAGTGACACACCGGTAAAGTCAACCCATTCTTT	5131
QY	2393	ctgtgtatcccttgtagtggcccgtagccctctaggttagagacttaccgtaccacaatggg	2452
Db	5130	CCTCCATGTACCAATTTGCCCTTAATCCTCTGCTAGGAAAGATTTGCTACTCAAACTTAA	5071
QY	2453	agctcaaatcttcttga---acaagaagaccagaagctgtctgtaataacaacccat	2509
Db	5070	AGCCCAATTTACTTTTAAAGGATCAGAGGTGATGGAGACCAATGGACACCCCT	5011
QY	2510	cactgtgtgaacctccaattagatgatatcgaatatctccccaagtaagac	2569
Db	5010	GCAAGTCTGCTTAACCTTAACATATGACAGATGATGTCGGCTACATGAGACCTCAAAAGGSCC	4951
QY	2570	tgatcaagaatataaagc---tgytttgaagcagttccccaagcctggacaagaacgc	2626
Db	4950	AGATGTGCTCTTAAGGTTCACATGGCTCTGTGATTTTCCCAAGGCTTGGGCAAGAACCGG	4891
QY	2627	agggaatgggttgcacaagaagaatctcccaacaggttattcaactggaagccaagtgctac	2686
Db	4890	GGGATGGGGCTGGCGTTCGCCAACCTCTCTGATCATACCTCTGAAGGCAACCTCTAC	4831
QY	2687	accagatcacagcacagtaaccctctgtagtagagaagctcgagaagaatttggcgcga	2746
Db	4830	CCCCGTTCATTAACCAATTAACCCCATGTGCACAGAGGCCAAGCTGGGATCAACCCCA	4771
QY	2747	tgttcaagaattatccaagaaggaatcctagttcctgtgccaatcccttggaaataccc	2806
Db	4770	CATCAAGAGACTGCTGGATAGGGAAATTCGTGATCCCTGACGTCCTCCCTGGAAACACC	4711
QY	2807	ccctgcaccggttaggaagccttggagacaatgattatcgaccagtafacaggtacttgaaga	2866
Db	4710	CTCTCTACCGGTTAAGAAACCGGGGACTATATATTATAGCCTGTCCAGATCTGACAGA	4651
QY	2867	ggtcaataaagaaggtgcagagacatacaccaacacggtcccgaaacccctlaaacctcttga	2926
Db	4650	AGTCAACAAACGGGGTGAAGACATCCACCCCAACCGGCCCAACCTTACAACTCTTGAG	4591
QY	2927	cgccctccgctcgaaucgagactggttaaccagtatctgacttaagaatgcctctctg	2986
Db	4590	CGGGCTCCACCCGTCACACAGGTGATCACTGTGCTTAATAAAGATGCTTTTCTTG	4531
QY	2987	ccttagatlaaaccacactgcaaacacacttcttgcttcgtagtgaagaatccaggtac	3046
Db	4530	CTTAAGACTCCACCCCAACAGTGAAGTCTCTTTCGCTTTTGAGTGAAGATCAAGAGAT	4471
QY	3047	gggaagaacccgagctcaactcgagaccgagctgcaccaaggttccaagaactcccgac	3106

Db	4470	GGGATCTCAGAGCAATTAACTGTGACCAAGCTCCGCAAGGGTTTCAAAAACAGTCCAC	44111
QY	3107	gacttgaagaaagccctaacagagacccctgagccaaactcaagatccaacaccccaagt	31666
Db	4410	CCTGTTTATATAAAGCCCTGCACAGAGACCTTCGAGACTTCCGATCCAGCACCCAGACT	43511
QY	3167	gaccctccctccagtaacgtatgataccctgctctcgtgagagccacccaacagatcgtct	32266
Db	4350	GATTGCTGCCAGTATGAGATGACCTTACTGCTGGCCGCCACTTCTGACTTGACTGTGCA	42911
QY	3227	agaagtaagaaagcactactgctgtgaaattgtctgaacctagctacagagcctctgtaa	32866
Db	4290	ACAAAGTACGGGGCCCTGTGTACAAACCTTAGGGGACCTCGGATATCGGGCCCTCGCCAA	42311
QY	3287	gaagggcccaagatttcagagagagaggtgaataactcttggtgtacagtttgccggagggca	33466
Db	4230	GAAAACCCCAAAATTGGCCAGAAACAGCAAGTATCTGGGGTATCTCTTAAAGAGGGTCA	41711
QY	3347	gcgaatgctgcaggaagcagcaggaanaactltaglccagatataccggcccccaacacagc	34066
Db	4170	GAGATGGCTGACTAGGGCCAGAAAGAGACTGTGATGGGGCAGGCTACTCCGACACCC	41111
QY	3407	caaaacagtgagagatttttgggagcagctltgaaatttgcaactgtgatcccggtgt	34666
Db	4110	TGCAACAACTAAGGGAGTTCCTTAGGGAGCGGACGAGCTTCTCCCTCGATGATCCCTGGST	40511
QY	3467	tggaacctgaaagccccactcttaccgcgtcaacaaagaanaagggatctcctgagc	35266
Db	4050	TGCAAAATGGACACCCCTTGTAACCTCTCACAAAACGGGAGCTGTGTGAAGTG666	39911
QY	3527	tcctgaagacccaagaagcatltgatctatacaaaaagccctgctgagcgcaactgctct	35866
Db	3990	CCCAAGACCAAGAAAGGCGCTACCAAGAGATCAAGCAGCGCTCTTAATCGCCCTG66CT	39911
QY	3587	ggccctccctgaagtaactaaacoccttaacoccttaatgtgtgaaagcgtgaaggaagtagc	36466
Db	3930	GGGATTTGGCAACACTTGACTTAAGCCCTTCGAACTTTTGTGTGACGAAGAGCAGGCTACGC	38711
QY	3647	ccgagggagtttlaacccaaccccttagaacatctgagagagaaactlgttgcctactgtcaaa	37066
Db	3870	CAAAAGGTCTCTTAACGCAAAAACCTGGGGGCTGTGGCTCGGGCGGTGGCTACTCTGTCCA	38111
QY	3707	gaagcttatctcttgaacgaatggtgtgagccgcatgtctcgaagcatalcgcagcgtgtgc	37666
Db	3810	AAAGCTAAGCCCACTGGGCACTGGGTGGCCCTTGTGCTACAGGATGTGATGACGACATGCG	37511
QY	3767	calactgtgcaagagcgtgacaaatlygactlttgagagaaatataactgttaatlagcccc	38266
Db	3750	CGTTCTGCACAAAGACCTCTGGCAAGCTACCAATGAGACAGCAATGTCATTTGTGGCCC	36911
QY	3827	ccatgcatltgagagaactlgttcgcagagcccccagacgaaigtatgagcaaacagcccgat	38866
Db	3690	CCATGCA6TAAAGGCACTACTTAAAGCAACCCCTGTGATGCTGTGGCTCTCCAAAGCCCGAAT	36311
QY	3887	gaccacatataaagcctgtctct---caagagaaggtgcactcttgctccacaagccgcg	39431
Db	3630	GACCCACTACCAAGGCTTCTGTTGTGACACAGGACGAGTCAGCTTCGAGCCAAATATGT6GC	35711
QY	3944	ttctaacacctgcaactctcttcgctccttgaagaagatgataagacaagtacatltgatltgca	40031
Db	3570	CCTTAAACCCAGACTACGCTGCTGCCCTCA---CCTGAGAGGGGCTGTCAAAATGACTGCT	3514
QY	4004	tcaactatgtattgagagaagacttggtgtccgcaagaagccttaacagacataccgctgtactg	4063
Db	3513	TGACATCTTGCGTGAACCCCAACGAACTAGACCAAGATCTTATCGACGACAGCCTCTCCAGA	3454
QY	4064	agaagtgttaaccctggtctactcgaagaagcagatagtgtgtggaaggttaagagatagc	4123
Db	3453	CGCTGACCAACCTCGGTACCAAGATGGAGACCTTCTGTCAAGAGGGGCGACGCCAAGGC	3394
QY	4124	tgaggcgagcagttgttgaaggaagcccgcaacgaatcttggtgcagcgaagcctgcgcgaaggaac	4183

Db	3393	CGGAGCAGAGTAAACCCGAGACCGAGGAGTCTGGCCCAAGCACTGCCAGCCGGGAC	3334
Qy	4184	ttcagcgcaaaaggctgtagctcatgycctctcaagcaagctttgcygctcggccgaaggaa	4233
Db	3333	ATTCGCCCAAGAGGCTGTGATTTGATTCAGCTCCACCCAAAGCCTTTAAAATTTGGCAGAAAGTAA	3274
Qy	4244	atccataaacattatctacgcagcagaggtatctcccttgagctgacacgctacagagggc	4303
Db	3273	GAAGCTGAATGTTTACACCGATATGCGGTTATGCTTTTGGCACTGCCCATTTTCACGGGAGA	3214
Qy	4304	catctataacaaagggtgtgtcttacctcagcagggagggagataaagaacaaagaga	4363
Db	3213	AATATATAGAAAGGGCGGGTGTGCTCAATCAGAAAGGAAAGAAATCAAAAATTAAGAGAGA	3154
Qy	4364	aattctaagcctattagaagccttaccatttgcacaaaggctagctattatacactgtcc	4423
Db	3153	GATTTTGCCCTACTGTGAAGGCTCTCTTCCCTGCCCAAAAGCACTTACATTAATTCATTTGCC	3094
Qy	4424	tggscatctagaagaagccaaagatctcatctctagaggaaacagatggctcggccgggttgc	4483
Db	3093	GGGACATCAGAAAGGAAACC GCCGAGCAGCAAGGGGCAACAGATGGCGGACCAAGCGGC	3034
Qy	4484	caagcagcgacccccagcctgttctaacctctgcctctataatagaagaagccaaagcccca	4543
Db	3033	CCGAGAGTATGCGCCTAGAAAACTCCAGAGATTTCCACACTTTCGATGAGAAATTTACGC	2974
Qy	4544	aacc-agaagcagtagacacctagaagaactgycagagataaanaagatagaccagttct	4602
Db	2973	CCCCATATACATGATGAACATTTTTCATATACGGGATCGACATGAAAGATTCCTGACTAACT	2914
Qy	4603	ctgagactcgcgagggagacc-----tgcataactctatatgggaaggaatctctcc	4654
Db	2913	AGGGGCCACTTATGACGATCCAAAGAGTGTGGTTTTCAGGGGAAACCTCTGATATCC	2854
Qy	4655	ccacaagaagaggttagaataitgccaaagatatacgtlctaacccactagaagactaa	4714
Db	2853	TGATCATTAATCACTTTGAATATTAGATTTTCTTATCATTAATGACCCACTCATGTTCTC	2794
Qy	4715	acaccgtgcagcagttgtgtcagaaca-----tccctatcatgtctcgaagctcag	4768
Db	2793	AAAAACAAGGCTCTTCTTACAAAAGAACTACTGTCTCTTATTCATGATGTGAACCCGGATCG	2734
Qy	4769	agtgctgcactcggttgtcaacatitgttgcctctcagcctggtttaatgctatcttc	4828
Db	2733	AAGCCTCAAAAGCATCTCAGAGACTTGCACAGCCTTGTGCACAGGTCAATGCCAGCAATGC	2674
Qy	4829	cagatatcctccagaagaagactaaaggaggaagccacccacagcgcctcatctggaagtga	4888
Db	2673	TGCGGTCAAAACAAGGAGCTAGATTTGAGAGGACCGCACCGGACCCACTGGGAAATTGA	2614
Qy	4889	cttcactgaggtlaaagccgcgtctaatacgcgnaacaatactatgtgttttgttagcac	4948
Db	2613	TTTTCACTGAGGTAAACCTGCGCTGTATGGGTATTAATTCCTTTAGTTTTCATGTGACAC	2554
Qy	4949	ctttcagagataggtagagctctatctctctactaagaagaagacttcaaaccgttgtgtcaa	5008
Db	2553	TTTTCTCGAGGTGTAAACCTTTTCCACAAAGAAAGAACTGCCAANAAGTTGTACCAA	2494
Qy	5009	gaaataactggaagaaattttccaaagatttgaaataccaaagtaatagatagggtacagaa	5068
Db	2493	GAAGGTACTAGAAAGAAATTTTCCCGAGATTCCGCAATGCCACAGATTTGGGAACCGACAA	2434
Qy	5069	tgttcacagcttggttgcagagtaagtcagggactcgcgcaagatatvgggatgttatg	5128
Db	2433	TGGGCTGCGCTTCTCTCCAAAGGTAAAGTTCAGACAGTAGCCGATTAATTTGGGGGTTGATTTG	2374
Qy	5129	gaactgcatgtgtctatcagagcccaagctcaggacagcgttagagagagatgaatagac	5188
Db	2373	GAAACTACATTTGCTTTACAGACCCCAAGACTTATAGTTCAGGTAGAAAGATGAATAGGAC	2314
Qy	5189	cattaaagaagcccttaccaatgtgacacagagactgtgcaatgaatgttagtgcct	5248
Db	2313	AATCAGAGACCTTAACTAAATTTGACGCTTGCAACTGCGCTTAGGAGACTGGGCTCCT	2254

QY	5249	ccgcgccttgctgcttlltagaggtgaggaaccccttgacaggtttggagctgcagcccccta	5308
Db	2253	GCCTTCCCTTGGCCCTGTGATGGAACCCGCAACAGCGCGGGGCCCCCAAGGCTTCACCCCAATA	2194
QY	5309	tgaattgctctacggaggaaccccccggttgcgaagaattgctcttgacatagtgtctga	5368
Db	2193	TGAATCTTATATAGGGGACACCCCGCCGCTTGTAATCTTCCTGTATCCGTGACATGGCAAA	2134
QY	5369	tgtgtcgtcttcccaagccttltctcttaagtcctcaagcgtctcgatgtggttgagtgagcg	5428
Db	2133	GGTTACTCAATAAACCCCTCTCTCCCAAGCTCATTTTACAGGGCAGCTCTACCTGtGTCCACAGCA	2074
QY	5429	agcgtggaagcagctcccgaggagccctacacag-----gaggagacttgcgaatttcaca	5482
Db	2073	AGCTGTGAAGACCGTTGGCGGCACTTTACCAAGACACTAGACCGCGCGTATGTACTCA	2014
QY	5483	tcgcttccaagttgagatcaagtlcaltbtltagacgcacagctgacggaanaacctcgagac	5542
Db	2013	CCCTTTCCGGGTGGGTGACACAGTGTGGGTCCGACAGCACCAACTTAATAATCTGTAAGC	1954
QY	5543	tcggttgaagaggacattatctgttacttttagacacacaagcgttggaaagtctgaag	5602
Db	1953	CCGCTGGAAGAGACCCCTATACCGCTCCTACTAGACTACCCCAACGCTCTCAAGATGACGG	1894
QY	5603	aatcccaacctgaatcaltatcattcccaqtltaagccgcg-----gcacc	5647
Db	1893	CATTGCACAGCGTGAATCCACCTCCCAACGTAAAGCTGCCGACACAGATTTGACCCAC	1834
QY	5648	tcccgcactcggggtlbgagagcgaanaagactgaaatcccttgaagcttcgcctccatcg	5707
Db	1833	AGCAGAAATCGACATGGCGGTTCACAGCTCCCAAAATCCCTTAAGATTAAGATTGACCCG	1774
QY	5708	cctgttcttactacttaacaataactccccagcgacgtagtataacgacctatagaagact	5767
Db	1773	CGGAGCTCCTAATCCCTTAATTCTTCTCTCTCTCAAAAGGGCCAGATCCGACGA	1714
QY	5768	cgaaccccga	5777
Db	1713	CCCGGGCTCA	1704
RESULT 15			
US-09-111-085-1			
: Sequence 1, Application US/09111085			
: Patent No. 6100034			
: GENERAL INFORMATION:			
: APPLICANT: Stoye, Jonathan P			
: APPLICANT: Weiss, Robin A			
: TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope			
: FILE REFERENCE: 4238/75168			
: CURRENT APPLICATION NUMBER: US/09/111,085			
: CURRENT FILING DATE: 1998-07-07			
: EARLIER APPLICATION NUMBER: GB 9710154.7			
: EARLIER FILING DATE: 1997-05-16			
: NUMBER OF SEQ ID NOS: 16			
: SOFTWARE: PatentIn Ver. 2.0			
: SEQ ID NO 1			
: LENGTH: 2462			
: TYPE: DNA			
: ORGANISM: Porcine retrovirus			
US-09-111-085-1			

Query Match	19.1%	Score 1568.4	DB 3	Length 2462
Best Local Similarity	79.6%	Pred. No. 0		
Matches 2002	Conservative	0	Mismatches 396	Indels 116
				Gaps 8
QY	5410	tcgaatggtgtagcagcagcgctgtgaaagcaatccggagagcctactaagaaggaact	5469	
Db	1	tcgaatggtgtagcagcagcagctgtgaaagcaatccggagagcctactaagaaggaact	60	

QY	5470	tgcaagttccacatcgcttccaaagtctggagatcagtcatagtttagacgcacgctgacg	5525
Db	61	tgcagattccacatcgcttccaaagtctggagatcagtcatagtttagacgcacgctgacg	120
QY	5530	gaacactcgaaactcggcttgagaggagacctatctctgtaactttgaccacaacaaagctgtg	5588
Db	121	gaacactcgaaactcggcttgagaggagacctatctctgtaactttgaccacaacaaagctgtg	180
QY	5590	tgaagctgcaagagatcccccacctcgatccatgcatccacgcttaagccggcgccacctc	5649
Db	181	tgaagctgcaagagatcccccacctcgatccatgcatccacgcttaagccggcgccacctc	240
QY	5650	ccgactcggggtgagagccgaaagaactgagaatcccttaagcttcgcctcatcgcc	5709
Db	241	ccgactcggggtgagagccgaaagaactgagaatcccttaagcttcgcctcatcgcc	300
QY	5710	tgtttcccttactctaaanaaactccccagcgcaagtagtaaacgcttatagacactcg	5765
Db	301	tgtttcccttactctgtaanaaactccccagcgcaagtagtaaacgcttatagacactcg	360
QY	5770	aaccccataagaccttatacccttactctgctgctgctatattgacacctgataacggctgtaact	5829
Db	361	aaccccataagaccttatacccttactctgctgctgctatattgacacctgataacggctgtaact	420
QY	5830	gtaaatagcactcagagtgctgctccttagagcaccctggtgacctgaaactgactgattctgcg	5889
Db	421	atlaacgactcacaaggagggactccctccttgaggacctggtgcctgaaattatgtctgcg	480
QY	5890	ctccgattgattaaaccccgctgtta-----aaagacacctcccaacctagtcogtagt	5943
Db	481	ctccgattgattaaaccccgctgttcaatgaccagcgacaccccccgaattgactcogtgct	540
QY	5944	tatggtctctaattgctcccccagg---cacagagaagaagaatactgctgggggtcttcgg	6000
Db	541	tacgggttttaagctttgcccagagaccccacaataagaagaattatgtggaatccctcag	600
QY	6001	gaatcctctctgtagagatgagagctgcgtccactcccaacgatgagactggaatagtcgg	6060
Db	601	gattctctttgcaagcatatgagctgcatataacttctaagtagtggaatttggaatagtcga	660
QY	6061	atcctctccagagccgggtaaaattctctcttgttgaatcc-----6102	
Db	661	gtccctcagcagaagacagaatgaattactcttctgtttaaacaactcctacagttaatcoaa	720
QY	6103	-----ggcccggcgaagtacaatgagaaact	6131
Db	721	ttatattatgycatgvggagatgtagaaagttggtccaacgcygggtacaanaaagaagtg- acg	779
QY	6132	ataataagataagagctgctcccatcagacttaagatatactcaagaataaagttctacgt-	6190
Db	780	aaataaagcaataagactgctcatctcgcttagaactagattataaataaagtttccactgta	839
QY	6191	-aaagagaaacaggaataattccaagaatgagataaagtgtatagctgvgggaatagttt	6249
Db	840	aaaagagaaacaggaataattccaagaatgagataaagtgtatagctgvgggaatagttt	899
QY	6250	ttac-----tatatgvcggvggagcgagcggtccacttaacoccttgccttagatagaagac	6305
Db	900	ctatgvggagctctcvggagagaagaagaagactcgttccgcctacttcgcccacgaatagaac	959
QY	6306	gggagacaaacccctctgtgcaatgtaggacccgataaagttactgctgtgacaaggggccccc	6365
Db	960	tcagatgtgaacctcccggttgcctatagaccacaataaggtttgvcgcgaacaagaactccc	1019
QY	6366	ggcccttgagatccacgcgcatlaacttgcgggtgcccaatlaacctgcgtgcgcgtcgaat	6425
Db	1020	aatccaagaaca-----gaagccatctcc	1043
QY	6426	aacacagccgctcagcacaagataccatgtagttgattctctcaacaacagccttagaaactc	6485
Db	1044	taacccctctgattacaataacatacaactccttgagatcagttcccaactga-----	1088
6486	ccaagtgctctgttlaagacagagaaagactcttgaagttcatccagggagcttcca	6545	

Db	1089	gcttaacatcactatataaacaagggvcgaaacttcttaagctccatccagaagacgtttca	1148
Qy	6546	agccatcaactccacgcgacccctgatagccacttctctgtgtgcttgcatacctcaag	6605
Db	1149	agctctaactccacgactccacgagcgtaactctctctgtgtgcttgcatacctcgg	1208
Qy	6606	gctccttattatgtggggatgctgctaaagaagaataatcaatcgtgacccaagaagcatag	6665
Db	1209	cccacacttactatgtagggaaatcgtgcataaggaggaataatccaatgtacaaaagaacatag	1268
Qy	6666	aaatcaatgtatcatvrggggtccgcgaataagcttaccctacacgaagttccgggaagg	6722
Db	1269	agacaacatgtaacatvrgggaatcccaaaataagcttaacccctacgaggttctcvgaaaag	1329
Qy	6726	gacatgcatatgaaaagcttcccccatacccccacacaccttctgtatagctaatgtgtta	6783
Db	1329	caacgtgcataagatggttcccccataccccaacacaccttctgaaccacaacgaagcctt	1388
Qy	6786	tgagcaagcctccagaanaatacagatattagtaacctggttalaacagvtgvtgvgcatgca	6845
Db	1389	taatcgaaacactctvgagatcataatctcgttaacctgttatgacagvtgvtgvgcatgta	1448
Qy	6846	tacttggttaaccccccgtgtgtttccacctgaagcttccaaccaatccaaagatttctggt	6905
Db	1449	tactgtgattaaacccctgtgttctccacctgtgtttccaaccaaccaataagacttttggt	1508
Qy	6906	catgttccaatctgtcccccagatgataccaatcccgaggaagvtgvtcccttgatgata	6965
Db	1509	tatgttccaatgttcccccgggtgtatctactatcccgaaaagaacagtccttgatgata	1568
Qy	6966	tgactatctgataaaccgaaccaaagaagaccgataccctacacctagactgttaatgtc	7025
Db	1569	tgactatagatataatcgccaaaagaagcccatatccctgaacactagctgttaatgtc	1628
Qy	7026	cggtattagagccgcgttggctgtagaaacaagggaacagctgtccctgatacagaagca	7085
Db	1629	cggtattgtagatggtcgtccagcgvtgvgaaacaagaaacgctgtcccttaatacagaagccga	1688
Qy	7086	gacgtatagaaaagaagacttggctgtgacatactgvcgcacgaagaagatctccggacctt	7145
Db	1689	acagcttgaagaagaagactttagtaacctacatcgaattgttaacoggaagatctccaaagcctt	1748
Qy	7146	aaagatgcttcttagcaacctagaagaagtcctcgacattcttgtctgaagtgtgtctaca	7205
Db	1749	agaaaatactgtcagtaacctgtagagaatccctcaacctcttatctgaaagtgttctaca	1808
Qy	7206	gaaccgtaggggattagatctgctgttcttaagagaagvtggttatgtgcaagccttaa	7265
Db	1809	gaaccagaagggtttagatctgttatcttcaataagaagaaggttatgtgtagccttaaa	1868
Qy	7266	agaagaatgttgcttcatgtatagataactacagaagccatcaagaagctccaagaagact	7325
Db	1869	agaagaatgctgtctctcatgtatagataactacagaagccatcaagaagctccaagaagact	1928
Qy	7326	tagaaaaagatlaagagvggtctgaagggaaagaagvggtgcacaggggtgtgttgaag	7385
Db	1929	tagagaaaagatlaagagvggtctgaagggaaagaagvggtgcacaggggtgtgttgaag	1988
Qy	7386	atggttcaacaggtctcctctgtagtgaacccaacctgtcttctgtctgtagagggcccttagt	7445
Db	1989	atggttcaacaggtctcctctgtagtgaacccaacctgtcttctgtctgtagagggcccttagt	2048
Qy	7446	agtcctgtctctgttacttcaatgttvggctcgttgcttaataatagtttgttgccttgt	7505
Db	2049	agtcctgtctctgttacttcaatgttvggctcgttgcttaataatagtttgttgccttgt	2108
Qy	7506	tagagaaacgtagtgcatactcacaatactatgttactttaggcacaagtaaccaagccttc	7565
Db	2109	tagagaaacgtagtgcatactcacaagtaactgttactttaggcacaagtaaccaagccttc	2168
Qy	7566	gagccaagsgaaacttgaccttagaccttccagttcttaagattagaactttaacaaga	7625

Db 2169 gagccaagsgaagaaactgacctagccttccagcttctaagattagaactattaacaga 2228
Qy 7626 caagaagtgggaatgaataaagatgaataatgcaacctaaccttccagaacccaggaagt 7685
Db 2229 caagaagtgggaatgaataaagatgaataatgcaacctaaccttccagaacccaggaagt 2288
Qy 7686 aataaaagctctaataatgccccgaattacagaccctgtgctgccaagtaaatagtag 7745
Db 2289 aataaaagctctaataatgccccgaattacagaccctgtgctgccaagtaaatagtag 2348
Qy 7746 aaggtcacacttccattgttccaggcctgtatccctggcctaagtaagataacagaa 7805
Db 2349 aaggtcacacttccattgttccaggcctgtatccctggcctaagtaagataacagaa 2408
Qy 7806 atgagtgactaatcgcttattctggaattctgtaaaactgactggcaccataga 7859
Db 2409 atgagtgactaatcgcttattctggaattctgtaaaactgactggcaccataga 2462

Search completed: February 24, 2002, 02:58:34
Job time: 21265 sec


```

51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaIle 67
748 CCATCAGAGGAGGACCTTAAATTCCTGAATATCCCTGGCTGTTAAAGCCAAAT 797
67 eilIepheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84
798 CATTTTACAGACTGAGACCGGCTCTCATCTCTATCAGAGGCCCTTATATCC 847
84 eutThrTrpGlnAspLeuAlaGluAspProProTrrValLysProTrr 100
848 TTACGTCGAGAAATTTGGCAGAAAGATCCCTCCCATGGGTAAACCATG 897
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGlu 117
898 CTAATAAACAAGCAAGAGCCAGCTCCCAATCTGCTGGCTTGAAGAGAA 947
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
948 AAACAACAACCTCGCCGAAAGATGAGCCCTC.TCCTGTAATCTACCC 996
134 rAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
997 GAGATCGAGAGACCGCCGACCTGGCCGAGACCCCAACCTGTTCCCCACC 1046
151 ProLeuSerSerThrGlyCysCysGluGlyThr.SerAlaProProGlyA 167
1047 CCTTATCAGACAGAGGCTGTGAGGAGACCTCTCCCTCCTCTGAG 1096
167 lAProValValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
1097 CTCCTGGTGTGAGGAGGACCTCTGCGGAGCTCGAGCGGAGAGGCC 1146
184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTrrGly 200
1147 ACCCGGAGCGGACAGAGAGATCCGATATTTACCGCTGCGACCTATGG 1196
200 yProBrometProGlyGlyGlnLeuGlnProLeuGlnTrrTrrProPheS 217
1197 CCTCCCATGCGAGGCGCAATGACGCCCTCCAGATATGGCCCTTTT 1246
217 eSerAlaAspLeuTrrAspTrrLysThrAsnHisProProPheSerGlu 233
1247 CTTTGCAGATCTCTATATATGGAACCTAACCTCCCTCTCTCCGAG 1296
234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
1297 GATCCCAAGACGCTCAGCGGCTGTGAGGCTCTTATGTTCTCTACCA 1346
250 nProThrTrrPaspAspCysGlnGlnLeuGlnThrLeuPheThrThrg 267
1347 GCTTACTTGGATGATGTTCAACAGCTGTCAGACACTCTTCACAAACG 1396
267 lGluAlaGluArgIleLeuLeuGlnAlaArgLysAsnValProGlyAla 283
1397 AGAGGCGAGAGAGATCTCTTGAAGGCTTAAATAAATGTTCTCGGGCC 1446
284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePole 300
1447 GAGCGCGAGACCCAGCAGTTCGAAATGAGATTGACATGGGATTTCCCT 1496
300 uThrArgProGlyTrrPaspTrrAsnThrAlaGluGlyArgGluSerLeu 317
1497 GACTCGCCCGGTTGGACTACAAACGCGCTGAAGGTAGGAGAGACTTGA 1546
317 yIleTrrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
1547 AATTCATATGCCAGGCTCTGCTGGCGGCTCTCGGGGCGCTTCACAGCG 1596
334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluP 350
1597 CCCACTAATTTGGCTAAGGTAAAGAGGTGATCCAGAGGCCGACCAAC 1646
350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367

```

```

1647 TCCCTCGGTATTTCTTGAGAGGCTCATGAGACCTTCAGGGGCTTACCC 1696
367 ropPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
1697 CTTTGTATCCCTACAGAGGCCCGAAGAGCTCAGTGGCCCTGCTCC 1746
384 lIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 400
1747 ATTGGGCGAGTGGCTCTGGATATCAGAGAGAACTTCAGAGACTGAGAG 1796
400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluValT 417
1797 GTTACAGAGAGGCTGATGTTACGTGATCTAGTGAAGAGAGAGAGAGTGT 1846
417 yTrrArgArgGluThrGluGluGluGluGluGluGluGluGluGlu 433
1847 ATTACAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1896
434 ArgGluGluArgGluGluArgArgAspArgArgGlnGluLysAsnLeuTh 450
1897 AGAGAAAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1946
450 rLysIleLeuAlaAlaValAlaGluGlyLysSerSerArgGluArgGlu 467
1947 TAAGATCTTGGCCGCACTGTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1996
467 rAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuGly 483
1997 GAGATTTTAGAAATTTAGTCAAGGCTTACAGAGAGAGAGAGAGAGAG 2046
484 AsnArgThrProLeuAspLysAspGlnCysAlaTrrCysLysGluLysG 500
2047 AATAGGACCCCACTCGACAAGAGACAGTGTGCTATTTGAAGAAAAAG 2096
500 yHisTrrAlaArgAsnCysProLysGlyLysAsnLysGlyProLysVal 517
2097 ACCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2146
517 eAlaLeuGluGluAspLysAsp 524
2147 TAGCTTAGAGAGAGATTAAGAT 2169

```

seq_name: /cgn2_6/ptodata/2/ina/6b_COMB.seq:US-08-766-528-3

seq_documentation_block:

Sequence 3, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965
 REFERENCE/DOCKET NUMBER: MGP-038CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8132 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-766-528-3

alignment_scores: Length: 525
 Quality: 2640.00 Gaps: 1
 Ratio: 5.116
 Percent Similarity: 98.286 Percent Identity: 94.476

alignment_block:
 US-09-171-553B-4 x US-08-766-528-3 ..

Align seg 1/1 to: US-08-766-528-3 from: 1 to: 8132

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
 585 ATGGGACAGACGGGTGACGACCCCTTACTTACTCTGACCATTTGGAC 634
 17 rgluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
 635 TGAAGTTAAATCCAGGGCTCATTAATTGTCAGTTCCAGTTAAGAGGAC 684
 34 rotrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 685 CTGGGAGACTTCTGTGCTCTGATGAGCCGACATTCGATGTTGATGG 734
 51 ProSerGluGlyThrPheAsnSerGluLeuLeuAlaValLysAlaI 67
 735 CCATTCAGAGGGACCTTAAATTCGAGATTATCTGGCTGTAAAGCAAGT 784
 67 eilePheGlnThrGlyProGlySerHisProAspGlnIuproTyrlleL 84
 785 TATTTTTCAGACTGGAACCGGCTCTCATCCGATCAGAGACCTATATCC 834
 84 eutHrTrpGlnAspLeuAlaGluAspProProtrPValLysProTrp 100
 835 TTACGTGGCAAGATTGTGGCAGAGAGATCTCCGCGATGGGTTAAACCATGG 884
 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLul 117
 885 CTGAATAGCCAGAAAGCAAGGATCCCGAATTCGTGCTTGGAGAGAA 934
 117 sasnLysHisSerAlaGluLysValGluProSerSerSerTyrlleuPro 134
 935 AAACAAACACTCGGCTGAAAGATCAAGCCCTC.TCCTCATATCTACCCC 983
 134 rgsAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
 984 GAGATTAGAGAGCCACCGGCTTGCCGGAACCCCAATCTGTCCCCACCC 1033
 151 ProLeuSerSerThrGlyCysGlyGluGlyThr.SerAlaProProGlyAla 167
 1034 CCTTATCTGGACACAGGGTGGCCGAGAGGGACCTTGGCCCTCTGGAG 1083
 167 lAprOvalValGluGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
 1084 CTCCGCGCGTGGAGGAGCTCTGACAGGACTCGAGCGGAGCGGAGGGGCC 1133
 184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgTrpTYG 200
 1134 ACCCGGAGCGGACAGACGAGATCCGACATTCACCGTGGCCACGATACGG 1183

200 yProPromeLProGlyGlyGlnLeuGlnProLeuGlnIlyrTrpProPheS 217
 1184 CCTCCACACCGGGGGGCCCAATTCGAGCCCTCCAGTATTTGGCCCTTTT 1233
 217 erSerAlaAspLeuTyraSnrTrpLysThrAsnHisProProPheSerGlu 233
 1234 CTCTCGAGATCTCTATTAATTTGAAAACTAACCATCCCTTCTCTGGAG 1283
 234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
 1284 GATCCCAACGCGCTCACGGGGTGTGTGATCCCTTATGTCTCTACCA 1333
 250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThrg 267
 1334 GCCTACTTGGGATGATTGTCAACAGCTGCTGACAGACCTCTTCACACCG 1383
 267 lugiuarArgGluArgIleLeuLeuGlnLuarArgLysAsnValProGlyAla 283
 1384 AGGAGCGAGAGAGAAATCTATAGAGGCTAGAAAAATGTTCTCTGGGGCC 1433
 284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProle 300
 1434 GACGGGCGACCCACGCGGTTCCAAAATGAGATTGCATGGGATTTCCCTT 1483
 300 uThrArgProGlyTrpAspTrpAsnThrAlaGluGlyArgGluSerLeu 317
 1484 AACTCGCCCGGCTTGGGACTACAAACAGCGCTGAAGAGTAGGAGAGCTTGA 1533
 317 yslIeTyArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
 1534 AAATCTATCGCCAGGCTGTGGTGGGGGTCTCCGGGGCGCTTCAAGACGG 1583
 334 ProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPr 350
 1584 CCACATATTTGGCTACAGTAAGAGAAAGTATGACAGGACCGAATGTAACC 1633
 350 oProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrP 367
 1634 CCCCTCTGTTTTTCTTGAAGAGCTCTGTAAGCCTTCAAGGGGGTACACC 1683
 367 roPheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAlaPhe 383
 1684 CTTTGTATCCCACTCAGAGGCCCAAAAGCCTCAGTGGCTTGGCCCTTT 1733
 384 lIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 400
 1734 ATAGGACAGTCAAGCTTGATATTAGAAAGAAAGCTTCAAGAGCTGAAGG 1783
 400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValT 417
 1784 GTTACAGAGGCTGAGTTACGTGATCTAGTGAAGAGGACGAGAAAGTAT 1833
 417 yTrpArgArgGluThrGlnGluGluLysGluGlnArgLysGluLysGlu 433
 1834 ATTACAAAGGAGACAGAGAAAGAAAGGAGCAAAAGAAAGAGAGAGA 1883
 434 ArgGluGluArgGluLuarArgAspArgGlnGlnGluLysAsnLeuTh 450
 1884 AGAGAGAAAGGAGGAGAAAGACGTATTAACGCGCAAGACAAATTTGAC 1933
 450 rLysIleLeuAlaAlaValAlaGluLysSerSerArgGluArgGluAla 467
 1934 TAAATCTTGTGGCTGCGAGTGTCAAGGAAAGCAATACGAAAGAGAGA 1983
 467 rgsAspPheArgLysIleArgSerGlyProArgGlnSerLysAsnLeuGly 483
 1984 GAGATTTTAGGAAATTAGGTACGGCCCTAGACAGTCAAGGAAACCTGGGC 2033
 484 AsnArgThrProLeuAspLysAspGlnCysAlaTyrcysLysGluLysG 500
 2034 AATAGGACCCCACTCGACAGACGACATGTGCTATTTGTAAGAAAGAG 2083
 500 yHisTrpAlaArgAsnCysProLysGlyLysGluLysGlyProLysValL 517

```

|||||
2084 AACTGCGGCAAGCACTGCCCAAGAGGAAACAAAGACCAAGATCC 2133
517 euAlaLeuGluGluAspLysAsp 524
|||||
2134 TAGCTCTAGAACAGATTAAGAT 2156

```

seq_name: /cgn2_6/plodata/2/ina/bh_comb.seq:us-08-766-528-1

seq_documentation_block:

```

: Sequence 1, Application US/08766528
: Patent No. 6190861
: GENERAL INFORMATION:
: APPLICANT: Jay A. Fishman
: TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
: NUMBER OF INVENTIONS: AND METHODS OF USE
: NUMBER OF SEQUENCES: 74
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 60 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/766,528
: FILING DATE:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/572,645
: FILING DATE: 14-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Louis Myers
: REGISTRATION NUMBER: 35,965
: REFERENCE/DOCKET NUMBER: MGP-038CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8060 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-766-528-1

```

alignment_scores:

```

Quality: 2634.00 Length: 525
Ratio: 5.105 Gaps: 1
Percent Similarity: 98.286 Percent Identity: 94.286

```

alignment_block:

US-09-171-553b-4 x US-08-766-528-1 ..

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1 MetGlyGlnThrValThrThrProLeuSerLeuPheLeuAspHisTrpTh 17
|||||
3112 ATGGGAGACAGCGTGCACGACCCCTTGTAGTTGACTCTCGACCAATTGGAC 3161
17 rGlValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
|||||
3162 TGAAGTTAATCCAGGCTCATTAATTTCAGTTTCAGTTAAGAGGAGAC 3211
34 rGTrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
|||||
3212 CTTGGCAGACTTCTGTGTCTGTGAATGGCCGACATTCGATGTTGGATGG 3261

```

```

51 ProSerGluGlyThrPheAsnSerGluLeuLeuAlaValLysAlaIle 67
|||||
3262 CCATCAGAGGGGACCTTAATTCTGAGATTATCTGCGCTGTTAAGACAT 3311
67 eLLePheGlnPheGlyProGlySerHisProAspGlnGluProTyrIle 84
|||||
3312 TATTTTACAGACTGACCCGGCTCTCATCCGATCAGACGCCCTATATCC 3361
84 euThrTrpGlnAspLeuAlaGluAspProProTyrValLysProTyr 100
|||||
3362 TTACGTGCAGAGATTGGCAGAGATCTCCGCAATGGTTAAACATATG 3411
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLys 117
|||||
3412 CTGATTAAGCCCAAGAAAGCCAGGCCCGCAATCTGCTTGAGAGAA 3461
117 sAsnLysHisSerAlaGluLysValGluProSerSerTyrLeuPro 134
|||||
3462 AAACAACACTCGGCTGAAAAAGTCAGACCTC.TCCTCATATCTACCC 3510
134 rGAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
|||||
3511 GAGATTAGAGAACACCGCCTTGCGGAAACCCCAATGCTGCCCCAC 3560
151 ProLeuSerThrGlyCysCysGlyGlyThr.SerAlaProProGly 167
|||||
3561 CCTTATCTGACACAGGAGTCCCGGAGGGGACCTTTGGCCCTCGGAG 3610
167 LaProValAlaGluGlyProAlaAlaGlyThrArgSerArgGlyAla 183
|||||
3611 CTCGGGCGTGGAGGACCTTCTGAGGGAGCTGGAGCGGAGGGCGCC 3660
184 ThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrGly 200
|||||
3661 ACCCGGAGCGGACAGACGATCGGACATTAACGCTCGCACCTACGG 3710
200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPhe 217
|||||
3711 CCTCCACACCGGGGGGCCAATTGCAGCCCTCCAGTATGCGCCCTTT 3760
217 eSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
|||||
3761 CTCTCTCAGATCTCTATTAATTGAAAACATTAACATCCCTTCTCGGAG 3810
234 AspProGluArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
|||||
3811 GATCCCAAGCGCTCACGGGGTGGAGTCCCTTATGTTCTCTCACCA 3860
250 nProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 267
|||||
3861 GCTACTTGGAGATGATGTCACACAGCTGCTGCAGACATCTTCCACACCG 3910
267 LuGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAla 283
|||||
3911 AGAGCGAGAGAGAAATCTATAGAGCTAGAAAAAATGTTCTCGGGGCC 3960
284 AspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300
|||||
3961 GACGGCGACCCACCGCGGTGGAAAAATGATGACATGGGATTTCCCTT 4010
300 uThrArgProGlyTyrAspTyrAsnThrAlaGluIleArgLysSerLeu 317
|||||
4011 AACTCGCCCGGTTGGACTACAAACAGCGCTGAAGATGAGGAGAGCTTGA 4060
4061 AAATTTATCGCCAGGCTCTGTGGCGGGTCTCCGCGCGCTCAAGACGG 4110
317 ySLeTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArg 333
|||||
334 ProThrAsnLeuAlaLysValArgLysValMetGlnGlyProAsnGlu 350
|||||
4111 CCACATATTTGGCTAAGTAAAGAAAGATGACAGGACGGAATGAACC 4160

```

```

350  opSoseValPheLeuLugLuArgLeuMetGluAlaPheArgRphPheRhp 367
      |||||
4161  CCCCTCTGTTTTTCTTGAGAGGCTCTTGGAAGGCTTCAGGGCGGTACACC 4210
367  robPheasProthrSerGluAlaGlnLysAlaSerValAlaLeuAlaPhe 383
      |||||
4211  CTTTGTATCCCACTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTT 4260
384  lIeGlnSerAlaLeuaspIleArgLysLysLeuGlnArgLeuGlnGlu 400
      |||||
4261  ATAGGACAGACACCCCTTGATATTAGAAACAAGCTTCAGAGACTGGAAG 4310
400  yLeuGlnGluAlaGluLeuArgaspLeuValArgGluAlaGluLysValT 417
      |||||
4311  GTTACAGGAGAGCGTAGATTACGTATCTAGTCAAGGAGCGACAGAAAGTAT 4360
417  yTfYrArgArgGluThrGlnGluGluLysGlnArgLysGlnLysGln 433
      |||||
4361  ATTACAAAAGGAGACAGAAAGAAAGGACAAAGAAAAGAAAGAGAGAA 4410
434  ArgGluGluArgGlnGluArgArgaspArgArgGlnGlnLysAsnLeuTh 450
      |||||
4411  AGAGAGAAAGGAGGAGAAAGACGTAATATAAGCGCAAGAGAAGATTGGAC 4460
450  rlySllLeuAlaAlaValValGluGlyLysSerSerArgLugArgGluA 467
      |||||
4461  TAAGATCTTGCGTCGAGTGGTGTGAAGGGAAGAACGTAACGGAAGAGAGA 4510
467  rGaspPheArgLysLleArgSerGlyProArgGlnSerGlyAsnLeuGly 483
      |||||
4511  GAGATTTTGAAGAAATTAGTCAAGGCCCTTGACAGTGAAGGAACTGGGC 4560
484  AsnArgThrProLeuaspLysaspGlnCysAlaThrCysLysGluLysG 500
      |||||
4561  AATAGAGACCCCACTCGACAAGAGCAATGTGCTATATTGTAAGAAGAGG 4610
500  yHlStrPalaArgAsnCysProLysLysGlyAsnLysGlyProLysValL 517
      |||||
4611  ACACCTGGGGAAGGAACTGCCCAAGAAAGGAAACAAAGGAGCAAGAGATCC 4660
517  euAlaLeuGlnGluAspLysasp 524
      |||||
4661  TACCTCTAGAGAGATTAAGAT 4683

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-075-272-1
seq_documentation_block:
: Seq. documentation 1, Application US/09075272
: Patent No. 6136598
: GENERAL INFORMATION:
: APPLICANT: MILLER, A. DUSTY
: APPLICANT: WOLGAMOT, GREG
: APPLICANT: BONHAM, LYNN
: TITLE OF INVENTION: MUS DUNNT ENDOGENOUS RETROVIRAL
: TITLE OF INVENTION: PACKAGING CELL LINES
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th floor
: CITY: San Francisco
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/075_272
: FILING DATE: 08-MAY-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
:

```

```

1  APPLICATION NUMBER: 09-MV-1997
2  FILING DATE: 03-MAY-1997
3  ATTORNEY/AGENT INFORMATION:
4
5  NAME: POOT, Brian W.
6  REGISTRATION NUMBER: 32,928
7  REFERENCE/DOCKET NUMBER: 14538A-003710
8  TELECOMMUNICATION INFORMATION:
9  TELEPHONE: (206) 467-9600
10 TELEFAX: (415) 576-0300
11 INFORMATION FOR SEQ. ID NO.: 1:
12
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 8655 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18
19 MOLECULE TYPE: cDNA
20
21 IS-09-075-272-1

```

[illegible]

```
alignment_block:
US-09-171-553B-4 x US-09-075-272-1 .
```

Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655

```

1 MetGlgLThrValIthrProLeuSerLeuThrLeuAspHisTyrPh 17
   |||||.....|||||.....
722 ATGGAGCAATCTGTCTCCACCTCCCTTCCTCTAAACCCTGGACATTGGA 771

17 rglUvaIArgSerArgAlaHisAsnLeuSerValGlnValLysIsglyP 34
   |||||.....|||||.....
772 GGAGTGAAAAAATCAGAGCCACAACAATCAGTCAGTAGAGTTAGAAAGGCC 821

34 rofPrpInThrPheCysAlaSerGluTrpProThrPheAspYalGlyTrp 50
   |||||.....|||||.....
822 CATGGCAAACCTTTTGCCTCGCATGGATGCCAACGTTGGAGTGGGCTGG 871

51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAla 67
   ||| |||||.....: : : : : |||||..... ||
872 CCACCGGAGGATGCTTTGACTTGCATCTGCATGATGCCGCCCTAGGCGAAT 921

67 eIllepEgInThrGlyProGlySerHisProaspGlnGluProTyrlle 84
   ::|||::|||
922 TGTTTTTCAGGAA...GAAGGGGGTCCACCTGATCAGATCCCTTAATGG 968

84 euThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
   :::|||||::|||::|||::|||::|||::|||::|||
969 TCACCTGGAGAGTCTCGTCAGATTCCACCTTGGTGGGTCAAGCCCTGG 101B

101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyIuLy 117
   |||||::||| |||||::|||
1019 ..... ACCCCAATACTCTCGAACAATGACGGTTCAGATTGC 1055

117 sAsnLysHisSerAlaGluLysValGluProSerSer....SerTYRL 132
   ::|||::|||::|||::|||::|||::|||::|||
1054 CCAATCCGATGTCAGTGAAMAATCTGTCGCCACACCCCCAGATT 1103B

132 euProArgAspArgGlyAlaIleAspLeuAlaGlyLThrProThrcySser 148
   ||||::||| ::|||
1104 ATCCAGAAGTTACGACCTCCTCTGGATGGATCCCAACCTCCCTTAC 1155B

149 ProThrProLeuSerSerThrGlycysCysGluGlyThrSerAlaPro.. 164
   ||| ||| ::|||::|||::|||::|||::|||
1154 CCCCTGCCCAACGACGACACTGTCACAGACCCCACTGTGGCAGGCCCTCA 1203B

165 .....ProGlyAlaProValAlaGluGlyProAlaIlaGlyThrArgS 179
   ||| ||| ::|||::|||::|||::|||::|||
1204 GCCGACAAACACAGGAGCGGGCTCAGAGACCGCGGAGGACTTCGGA 1253B
```

```

179 eArGArgGLyAlaThrProGluArg.....ThrAspLeuIleAla 192
|||||.....|
1254 GCCGCCGGGGTGGAGCCCTGCAGAGGAAGGGGGCCAGATTCCACACTT 1303
193 IleLeuProLeuArgThrTyr...GlyProPromeProGly...GlyGI 207
|||||.....|
1304 GCCCTTGCCCTTAGAGCTCATGTGGAGAGGCCAACGCCAGGACCTAATGA 1353
207 nLeuGIProLeuGIInTyrTriProPheSerSerAlaAspLeuTyrAsnT 224
|||||.....|
1354 TCTCATCTCTTACAGTACTGCTTCTTCTCTCTGATTATATATAT 1403
224 rPLySThAsnHisProPheSerSerGluAspProGluArgLeuThrGly 240
|||||.....|
1404 GGAAGAACCTACCCCTCTCTGAGAAACCTCTGGCTTACTGAGG 1453
241 LeuValGIUserIleuMetPheSerHisGlnProThrTyrPaspPysGI 257
|||||.....|
1454 CTCCTTGAAATCCTTATGTCTCTCATCAGCCCACTTGGATGATGCTCA 1503
257 nGIInLeuLeuGIInThrLeuPheThrThrGluGluArgGluArgIleuL 274
|||||.....|
1504 GCAGCTTTGGAGTTCTCTTCTTACACAGAGGAAGGAAGGAATCCTGA 1553
274 eugIAlaArgGlyAsnValProGlyAlaAspGlyArgProThrGluLeu 290
|||||.....|
1554 TGGAGGCAAGAAAAAGCTTCTAGAGAGGAGGACACCACTGCTCTC 1603
291 GlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTyrPaspTy 307
|||||.....|
1604 CCTAACCTCTGAGAGAGCTTTCCCTTGAACCCCTCAACTGGGACTA 1653
307 rAsnThrAlaGluArgGlyUserLeuLysIleTyrArgGlnAlaLeuV 324
|||||.....|
1654 CAACACTGGGAGAGTAGGGAGCGCTCTCTTCTTACCGTGGAGCTGG 1703
324 aAlaIleGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
|||||.....|
1704 TGGCAGGCTCTCAGAGAGCCGCAAGACGCCCAATTTGGCTAAGGTA 1753
341 ArgGluValAlaMetGlnGlyProAsnGluProPheSerValPheLeuGluAr 357
|||||.....|
1754 AGAGAGGCTTCTCAGGGGACAGCTAACCCCTCAGTCTTCTTGAGAGC 1803
357 GLeuMetGluAlaPheArgArgPheThrProPheAspProThrSerGluA 374
|||||.....|
1804 TCTTATGAGGCGGTATAGAGGTACACCCCTTTCATCCCTCGTCAGAG 1853
374 IaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAsp 390
|||||.....|
1854 GACAGAAAAGCCGCTGATGACCATGCTTATGGCAGCTGCTCCGAT 1903
391 IleArgLysLysLeuGlnArgLeuGlnGluLeuGlnGluAlaLeuAr 407
|||||.....|
1904 ATTAAGAAAAGCTACAGAGCTGSAAGGGCTCCCAAGATTATAGCTCCA 1953
407 gAspLeuValArgGluAlaGluLysValTyrTyrArgArgGlnThrGluG 424
|||||.....|
1954 ACATTATAGTGAAGAGAGAGAGAGAGTTATCACAAGAGAGAAACAGAG 2003
424 IuGIuLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg 440
|||||.....|
2004 AAGAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2053
441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaIleValVa 457
|||||.....|
2054 CGAGATCGCGCTCAGAGAGAGAAATTGAGATTGCGCGCGCTTAT 2103
457 IgLugLysSerSerArgGluArgGluArgAspPheArgLysIleArgS 474
|||||.....|
2104 AAATGATGAGCAGTCAGAAAAA..... 2125
474 eArgLysProArgGlnSerGlyAsnLeuGlyAsnArg..... 485

```

```

2126 .....GGCAGACAGATTCCTGGGCAACAGGGGACAGTAAACGCCCA 2167
486 .....ThrProLeuAspLysAspGlnCysAlaTyrCysLysGI 498
2168 GTGGCAGAAAGACCGCGCTGAGAAAAAGACCAATGCGCTTTGGCAAGA 2217
498 uLysGlyHisTirPalaArgAsnGlyProLysLysGlyAsnLysGlyProL 515
2218 AAAAGACACTGGGCTAAGAGCTGCCCTAAGAAA...AGAGGCAATTC 2264
515 yValLeuAlaLeuGluGluAsp 522
2265 AGGTCTGACCTAGAGAGAGAT 2287

```

```

seq_name: /cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-716-351A-1

```

```

seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Glibon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,351A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-128-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name="GalV SEATO Genome"
US-08-716-351A-1

```

```

alignment_scores:
Quality: 1660.50 Length: 542
Ratio: 3.944 Gaps: 10
Percent Similarity: 77.675 Percent Identity: 61.439

```

```

alignment_block:
US-09-171-553B-4 x US-08-716-351A-1

```

```

Align seq 1/1 to: US-08-716-351A-1 from: 1 to: 8535

```

```

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTyrPrh 17
|||||.....|
1001 ATGGGACAAAGATATTTCTACCCCTATCTCCCTCAGCTTAATATCAGTGGAG 1050

```

```

17  rgluValArgSerArgAlaHisAsnLeuSerValGluValLysLysGlyP 34
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1051 AGATGTGAGAACAGAGGGCTACAACTATCTCGTGAATCAAAAGGGA 1100
34  rotrpGlnThrPheGysAlaSerGluTrpProThrPheAspValGlyTrp 50
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1101 AATGGAGACTTCTGTCTCTCGAGTGGCCACATCGCGGTGGGGTGG 1150
51  ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaI 67
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1151 CCACGGAGGAGAACTTTAATCTCTCTCATTTTTCGATTAAAAAGAT 1200
67  ellePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1201 TGTCTTTCAGGAGAAC...GGGGGACATCCGGACCAAGTTCATATATCG 1247
84  eutrThrProGlnAspLeuAlaGluAspProProTyrValLysProTrp 100
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1248 TGGTATGGAGAGACCTCGCCGAGAAATCCGCCACCATGGGTG... 1288
101  LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGlu 117
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1289 .....CCAGCCTCCGGCAAGTGCCTGTGTCTCTGATAC 1323
117  sasnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuProA 134
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1324 CCGAGACCAAGTTCGGGGGAG...CCATCAGCTCTCCCGGACCC 1367
134  rGAspArgGlyAlaAlaasp.....LeuAlaGlyThrProThr... 146
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1368 CCATCTACCCGGCAACAGCACTTACTCTCTCTGAAACCCAGGCC 1417
147  .....CysSerProThrProLeuSerSerThrGlyCysCys 158
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1418 CCGCCCTATCCGGGGGCACTGCGACCCCTCTGGCCCTCAGGCATGG 1467
158  selGluGlyThrSerAlaProProGlyAlaProValGluGlyProAlaA 175
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1468 ACCGCCCTGCAGCAGATGCCGAGATAGCATGCTGAGGGCCAGCCG 1517
175  laGlyThrArgSerArgArgGlyAlaThrPro.....GluArgThr 188
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1518 CTGGAGACCAAGAGTCCGCGCCGAGTCCGACAGACACCTGGGTCT 1567
189  AspGluIleAlaIleLeuProLeuArgThrTyrGlyProProMetProG 205
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1568 GACTCCACTGTGATTTGGCCCTCGAGCAGATAGSACCCCGCGGACCC 1617
205  yelGlyLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAspLeuT 222
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1618 CAATGGCTGTGCTCTACAAATATTTGGCTTTTCTCAGCAGATCTTT 1667
222  yfAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeu 238
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1668 AATAATTGGAAATCTAATCATCCCTCTTTTTCGAAAAACCCAGAGTCTC 1717
239  ThrGlyLeuValGluSerLeuMetPheSerHisGlnProThrTrpAsp 255
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1718 ACGGGGCTCCTTGAGTCTCTATATGTTCCCATCAGCCCACTTGGAGCA 1767
255  pCysGlnGlnLeuLeuGlnThrLeuPheThrThrGluGluArgGluArg 272
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1768 TTGGCAACAGCTCCTACAGATTTCTTTCACCACTGAGGAGAGGAAAGAA 1817
272  leLeuLeuGlnAlaArgLysAsnValProGlyAlaAspGlyArgProThr 288
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1818 TTCTCTCTGGAGGCCGCAAAATGTCTTGGGAGACAATGGGGCCCTACA 1867
289  GlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTyr 305
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1868 CAGCTCGAAGAACCTTAAATAGAGGCTTCCCGCTCATATCAGCCACTG 1917
305  pAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnA 322

```

```

1918 GGATTACAAACAGCCGAGCTAGGAGACCTCTTCTGTGATACCGCCGA 1967
322  laLeuValAlaGlyLeuArgGlyAlaSerAlaGatGProThrAsnLeuAla 338
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
1968 CTTAGTGGCAGGTCTCAAAAGGGCAGCTCGGCTCTTACCAATTTGGCT 2017
339  LysValArgGluValMetGlnGlyProAsnGluProProSerValPheLe 355
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2018 AAGGTAAAGAGGCTCTTGAGAGGACCGGCAAGACCCCTTGCTTCTT 2067
355  uGluArgLeuMetGluAlaPheArgArgPheThrProPheAspProThr 372
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2068 AGAAGCCCTGATGAGGCGCTATAGAGATACACTCGTTGATCTCTT 2117
372  erGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGlnSerAla 388
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2118 CTGAGGGACAACAGGCTCGGCTCGCATGGCTTATTCGACAGTCAAGC 2167
389  LeuAspIleArgLysLysLeuGlnArgLeuGlnGlyLeuGlnGluAlaG 405
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2168 CCAGATATCAGAAATAAATTACAGAGCTAGAGGGCTCCAGGACTTTC 2217
405  uLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArgGluT 422
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2218 CTACAAAGATTTAGTAAAGAGGCAAAAGGTCTACCATTAAGAGAGA 2267
422  hrGluGluGluLysGluGlnArgLysGluLysGluValGluGluGluArg 438
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2268 CAGAAAGAAAGAAAGCAAGAAAGAAAGAAAGGAGGAGGAGAAAGAG 2317
439  GluArgArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaI 455
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2318 AGCGGGCGCGATAGCCGAGCAAAATAAATTGACTTAATTTGCGCCG 2367
455  aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2368 AGTAGTA.....AGTAGAGAA..... 2383
472  leArgSerGlyProArgGlnSerGlyAsnLeuGlyAsn..... 484
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2384 ..GGTCCACAGCTAGGCAAGAGCAAGCACTGACCAAGCAGGCAAGAG 2431
485  .....ArgThrProLeuAspLysAspGlnCysAlaTyrCys 496
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2432 ACACCTTAGGATGGAACACTCCACTAGACAAAGACCAAGTGGCATCTG 2481
496  sLysGluLysGlyHisTrpAlaArgAsnCysProLysLysGlyAsn...L 512
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2482 TAAAGAGAAAGGCGCATTTGGCAAGAGAAATGTCCTCCGAAAAAACAGCTCA 2531
512  ysgLysProLysValLeuAlaLeuGlu 520
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
2532 GAGAAAGCAAGGTTCTGAGCCCTAGAT 2557
seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq:us-09-011-745-3
seq_documentation_block:
; Sequence 3, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhito
; APPLICANT: Cosset, Francois-Loic
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; EARLIER FILING DATE: 1995-08-23

```



```

; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 7308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; US-09-011-745-3

alignment_scores:
    Quality: 1355.50      Length: 578
    Ratio: 3.440          Gaps: 15
    Percent Similarity: 68.166      Percent Identity: 48.962

alignment_block:
US-09-171-553B-4 x US-09-011-745-3 ..

Align seg 1/1 to: US-09-011-745-3 from: 1 to: 7308

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTPTTh 17
|||||
1567 ATGGCCAGACTGTTACCACTCCCTTAAGTTGAGCTTACGTCACCTGGA 1616
|||
17 rGlValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
|||||
1617 AGATGTCGAGCGGATCGCTCACACACAGCTCGTAGATGTCAGAGAAC 1666
|||
34 rOTrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
|||||
1667 GTTGGGTAACTTGTCTGTGCAAGATGCGCAACCTTAACTGCGAGTG 1716
|||
51 ProSerGlnGlyThrPheAsnSerGluLeuLeuAlaValLysAlaI 67
|||||
1717 CCGCAGAGCGACCTTAAACGAGACCTCAACACCGAGTTAAGATCAA 1766
|||
67 eIlePheGlnThrGlyProGlySerHisProAspGlnGluProGlyIle 84
|||||
1767 GGTCTTTTACCTGGCCGCGATGACACCCAGACAGGTCCCTCATGCG 1816
|||
84 eUThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
|||||
1817 TCACCTGGGAGCCTTGGCTTTTGAACCCCTCCCTGGGTCAAGCCCTT 1866
|||
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyL 117
|||||
1867 GTACAC...CTTACGCTCCGCTCT... 1890
|||
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTrpLeuPro 134
|||||
1891 .....CTTCTCCATCCGCGCCGCTCTCCGCC 1918
|||
134 rAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
|||||
1919 TTGAACCTCTCTGTCGACCCGCGCTCGATCTCCCTTTATCCACCCCTC 1968
|||
144 ThrProThrCysSerProThrPro.....LeuSerSerThrG 156
|||||
1969 ACTCCTTCTAGGCGCCAAACCTTAACCTCAAGTTCTTTCGACAGTGG 2018
|||
156 yCysCysGluGly.....ThSerAlaProPro... 165
|||||
2019 G.....GGGCGGCTCATGACCTACTTACAGAGACCCCGCGCTT 2059
|||
165 ..... 165

2060 ATAGGACCCAAAGACCCCTTCGACAGGAGCGAAATGGTGAGAA 2109
166 GAlaProValValGluGlyPro.....AlaAlaGlyTh 177
|||||
2110 GCGACCCCTTGGGAGAGGCGACCGAGACCCCTCCCAATGGCATCTCGCT 2159
|||||

177 rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIle 194
|||||
2160 ACGTGGGAGAGCGGAGACCCCTGTGGCCGACTCTCCACTTACCTGCGAGCAT 2209
|||
194 eUrProLeuArgThrTyGlyProProMetProGlyGlyGlnLeuGlnPro 210
|||||
2210 TCCCCCTCCCGCAGAGAGG.....AACGAGAG..... 2238
|||
211 LeuGlnIleTrpProPheSerSerAlaAspLeuTyAsnTrpLysThrs 227
|||||
2239 CTTCATACTGCGCGCTTCTCTCTCTGACTTGTACACTGGAATAATA 2288
|||
227 nHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGlu 244
|||||
2289 TAACCTCTCTTCTTCTGAAAGATCCAGGTAAACGACGACCTGATGAGGT 2338
|||
244 eUrLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
|||||
2339 CTGTTCATCACCACCATCAGCCCACTGGGAGCAGCTGTACAGCTGTGG 2388
|||
261 GlnThrLeuPheThrThrGluArgGluArgIleLeuLeuGluAlaAr 277
|||||
2389 GGGACTCTGCTGACCGGAGAAACAAACAGGCTGCTGTAGAGCTAG 2438
|||
277 gLysAsnValProGlyAlaAspGluArgProThrGlnLeuGlnAsnGlu 294
|||||
2439 AAGGCGGTGCGGCGCATATGCGCGCCCACTCACTGCCCCCAATGAAG 2488
|||
294 LeAspMetGlyPheProLeuThrArgProGlyTrpAspTrpAsnThrAla 310
|||||
2489 TCATGCGCGCTTTCCTCCCTGAGCGCCCACTGAGGATTACACACCGAG 2538
|||
311 GluGlyArgGluSerLeuLysIleTyArgGlnAlaLeuValAlaGly 327
|||||
2539 GCAGGTAGAAACCACTAGTCCACTATCCGCACTGCTCTTACCGGGTCT 2588
|||
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluVal 344
|||||
2589 CCAAAACGCGGCGAAGACCCCAACCAATTTGGCCAAAGTTAAAGAAATA 2638
|||
344 eTGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
|||||
2639 CACAAGGCGCCCATGAGTCTCCCTCGCGCTCTTACGAGACCTTAAGGAA 2688
|||
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
|||||
2689 GCGTATCGAGGTAACTCTTATGACCCCTGAGGACCCAGGCGCAAGAAAC 2738
|||
377 aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLys 394
|||||
2739 TAAATGTCTATGCTTTTCAATTGGCAGTCTGCCCCAGACATTGGGAGAA 2788
|||
394 yLsLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspLeuVal 410
|||||
2789 AGTTAGAGAGGTTTGAAGATTTAAACAAAGACGCTTGAGATFTGGT 2838
|||
411 ArgGluAlaGluLysValTyTyArgArgGluThrGluGluGluLysG 427
|||||
2839 AGAGAGCAGCAAAAGATCTTAAATAACGAGAAACCCCGGAGAAAGAGA 2888
|||
427 uGlnArgLysGluLysGluArgGluGluArgGluArgGluArg..... 440
|||||
2889 GGAACGTATCAGAGAGAAACAGGAGAAAGAAAGACGCGCTGAGCAGAG 2938
|||
441 .....ArgAspArgArgGlnGluLysAsnLeu 449
|||||
2939 AGGATGACGACAAAGAGAAAGAAAGATCTGAGAGACATGAGAGATG 2988
|||
450 ThrLysIleLeuAlaAlaValAlaGluGlyLysSerSerArgGluArg 466
|||||
2989 AGCAAGCATATTGGCCACTGTGCTGTAGTGACAGAGAAACAG..... 3027
|||||
```

```

466 uatgasphearglyslleargserglyproarglnserglyasnleug 483
      |||||:|||||
3028 .....CATGACAGGAGAGACAA...C 3046
483 lyaanargthrproleuasplyasplncysalatyrcyslysglyls 499
      :|||||:|||||:|||||:|||||
3047 GAAGAGAGTCCCAACTGATCGGACACAGTGTCTACTGCAAAAGAAAAG 3096
500 glyhistrpalaargasnrcysprolysglyls.....Glyasnlysglypr 514
      |||||:|||||:|||||:|||||:|||||
3097 GGCACCTGGGCTAAAGATGTGCCCAAGAAACACAGAGACTCTGGGGACC 3146
514 Olys.....ValleuAlaleuGluGlu 521
      |||||:|||||:|||||:|||||
3147 AAGACCCACGACCTCCCTCCCTGACGACCTAGATGAC 3180

seq_name: /cgn2.6/prodata/2/lna/6B.COMB.seq:US-09-011-745-4

seq_documentation_block:
Sequence 4, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weis, Robin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cosset, Francois-Joic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
EARLIER FILING DATE: 1998-06-22
PCT/GB96/02061
EARLIER APPLICATION NUMBER: 1996-08-23
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 7308
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-4

alignment_scores:
Quality: 1355.50      Length: 578
Ratio: 3.440          Gaps: 15
Percent Similarity: 68.166      Percent Identity: 48.962

alignment_block:
US-09-171-553B-4 x US-09-011-745-4 ..
Align seg 1/1 to: US-09-011-745-4 from: 1 to: 7308

1 MetGlnGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
|||||:|||||:|||||:|||||:|||||
1567 ATGGCCGAGCTGTACCACTCCCTTAAGTTGACCTTAGCTACCTGCGAA 1616
17 GlnValAlaArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
|||||:|||||:|||||:|||||:|||||
1617 AGATGTCCAGCGGATCGCTCAACCAAGTCGATGATGTCACAGAGAC 1666
34 roftrpGlnThrPhcysAlaSerGlnTrpProThrPhcAspValGlyTrp 50
|||||:|||||:|||||:|||||:|||||
1667 GTTGGGTAACTTCTGCTGCGAGATGGCCACACCTTTAACGTCGGATGG 1716
51 ProSerGlnGlyThrPhcAsnSerGlnLeuLeuAlaValLysAlaIle 67
|||||:|||||:|||||:|||||:|||||
1717 CCGCAGACGCGACCTTTAACCGAGACCTCAACCCAGGTTAAGATCAA 1766
67 eilepheGlnThrGlyProGlySerHisProAspGlnGluProTyrIleL 84

```

```

1767 GGTCTTTTACCTGGCCGATGACACCCAGACAGGTCCTCATCTG 1816
      :|||||:|||||:|||||:|||||:|||||
84 eutThrTrpGlnAspLeuAlaGluAspProProThrValLysProTrp 100
      :|||||:|||||:|||||:|||||:|||||
1817 TGACCTGGGAAGCCTTGGCTTTTGACCCCTCCCTGGGGTCAAGCCCTT 1866
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
      :|||||:|||||:|||||:|||||:|||||
1867 GTACAC...CCTAAGCTCCGCTCCT..... 1890
117 sasnLysHisSerAlaGluLysValGluProSerSerTyrLeuProA 134
      :|||||:|||||:|||||:|||||:|||||
1891 .....CTTCCTCCATCCGCGCCGCTCTCCCTCC 1918
134 rgaSp.....ArgGlyAlaAlaAspLeuAlaGly 143
      :|||||:|||||:|||||:|||||:|||||
1919 TTGAACCTCTCTGTTGACCCCGCTCGATCTCCCTTTATCCAGGCTTC 1968
144 ThrProThrCysSerProThrPro.....LeuSerSerThrGln 156
      :|||||:|||||:|||||:|||||:|||||
1969 ACTCCTTCTTAGGGGCAACCTAAACCTCAAGTCTTCTTGACAGTGG 2018
156 yCysCysGly.....ThSerAlaProPro.... 165
2019 G.....GGCGCGTCATGACCTACTTACAGAAAGACCCCGGCTT 2059
165 ..... 165
2060 ATAGGAGCCCAAGACACCCCTTCGACAGGACGGAATGTGGAGAA 2109
166 GlnAlaProValAlaGluGlyPro.....AlaAlaGlyTh 177
      :|||||:|||||:|||||:|||||:|||||
2110 CGGACCCCTGCGGAGAGGACCGGACCCCTCCCAATGATCTCGCT 2159
177 rArgSerArgArgGlyAlaThrProGluArgThrAspGlnIleAlaIle 194
      :|||||:|||||:|||||:|||||:|||||
2160 ACGTGGGAGACGGGAGCCCGCTGTGGCGACTACTCTCGCAGGCAT 2209
194 eutProLeuArgThrTyrGlyProProMetProGlyGlnLeuGlnPro 210
      :|||||:|||||:|||||:|||||:|||||
2210 TCCCCCTCCGCGAGAGAGA.....AACGACAG..... 2238
211 LeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysThrAs 227
      :|||||:|||||:|||||:|||||:|||||
2239 CTTCATATACGCGCGTCTCTCTCTCTGACCTTTACAACTGAAATAA 2288
227 nhisProProPheSerGlnAspProGlnArgLeuThrGlyLeuValGln 244
      :|||||:|||||:|||||:|||||:|||||
2289 TAACCTTCTTTTTCGAAATGATCCAGGTAAACTGACAGCTGTGATGAGT 2338
244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
      :|||||:|||||:|||||:|||||:|||||
2339 CTGTTTCATCACCACCATCACCCACCTGGGACGACCTGTGATGAGT 2388
261 GlnThrLeuPheThrThrGlnGluArgGluArgIleLeuLeuGluAla 277
      :|||||:|||||:|||||:|||||:|||||
2389 GGGACTGTGCTGACCGGAGAAAGAAACAGGGTGTCTTTAGAGCTAG 2438
277 GlnAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGln 294
      :|||||:|||||:|||||:|||||:|||||
2439 AAGGCGGTGCGGGCGATGATGGGCGCCCACTCACTCCCAATGAG 2488
294 leaSpMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnThrAla 310
      :|||||:|||||:|||||:|||||:|||||
2489 TCGATGCGGCTTTTCCCTCGAGGCGCAAGACGTGATACACACCCAG 2538
311 GlnGluArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLe 327
      :|||||:|||||:|||||:|||||:|||||
2539 CGAGAGCAACCAACCTAGTCCACTATGCCAGTGTCTCCAGGGGTCT 2588
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluVal 344
      :|||||:|||||:|||||:|||||:|||||

```

2589 CCAAAAGCGGCGAGAGCCCAATTTGGCCAGGTAAAGAAATAA 2638
 344 eGlnGluProAsnGluProSerValPheLeuGluArgLeuMetGlu 360
 2639 CCAAGAGGCGCCATATGATCTCCGCGCTTCTAGAGAGACTTAAGAA 2688
 361 AlapheArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
 2689 GCGTATCCGAGGTACACTCTTATGACCTGAGAGCCAGGAGGAAC 2738
 377 aserValAlaLeuAlaPheIleGlnSerAlaLeuAspIleArgLysL 394
 2739 TATGTGCTATGTCTTTCATTGCGCTGCGCCAGACATTGGAGAA 2788
 394 yLseuGluArgLeuGluGluLeuGluAlaLeuArgLysLysLysL 410
 2789 AGTAGAGAGGTAGAGATTAAACAGACGCTTGAGATTGGTGT 2838
 411 ArgGluAlaGluLysValTyrTyrArgArgGluThrGluGluLysG 427
 2839 AGAGAGGAGAGAAAGATCTTATTAACGAGAAACCCGAGAGAAAGAA 2888
 427 uGlnArgLysGluLysGluArgGluLysGluLysGluLysGluL 440
 2889 GGAACGTATCAGAGAGAGAAACAGAGAAAGAGAGAGAGAGAGAG 2938
 441ArgAspArgArgGlnGluLysLysLysLysLysLysL 449
 2939 AGGATGAGCAGAAAGAGAAAGAGAGATCGTAGAGACATAGAGAGAT 2988
 450 ThrLysIleLeuAlaAlaValAlaGluLysSerSerArgGluArgL 466
 2989 AGGACCTATTGGCCACTGCTGTAGTACAGAGAAACAG 3027
 466 uArgAspPheArgLysIleArgSerGluProArgGlnSerGlnLysL 483
 3028GATAGACAGGAGAGAGAA...C 3046
 483 LysAspArgThrProLeuAspLysAspGlnLysAlaTyrCysLysG 499
 3047 GAGAGAGGTCCCACTCGATCGCAGCAGAGTGTGCTACTGCAAGAAAG 3096
 500 GlyHisThrPalaArgAsnGlyProLysLysLysLysLysLysLys 514
 3097 GGGCACTGGCTAAAGATTGTCCCAAGAAACAGAGAGAGAGAGAGAG 3146
 514 Olys.....ValLeuAlaLeuGluGln 521
 3147 AAGACCCAGACCTCCCTCTGACCTGACCTAGATGAC 3180

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:us-09-011-745-2

seq_documentation_block:
 ; Sequence 2, Application US/09011745
 ; Patent No. 6165715
 ; GENERAL INFORMATION:
 ; APPLICANT: Collins, Mary KL
 ; APPLICANT: Weiss, Robin A
 ; APPLICANT: Takeuchi, Yasuhiro
 ; APPLICANT: Cosset, Francois-Joic
 ; TITLE OF INVENTION: Expression systems
 ; FILE REFERENCE: 09/011,745
 ; CURRENT APPLICATION NUMBER: US/09/011,745
 ; EARLIER FILING DATE: 1998-06-22
 ; EARLIER APPLICATION NUMBER: PCT/GB96/02061
 ; EARLIER FILING DATE: 1996-08-23
 ; EARLIER APPLICATION NUMBER: GB9517263.1
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: Patentl Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 7616
 ; TYPE: DNA

ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Portion of
 ; US-09-011-745-2

alignment_scores:

Quality: 1355.50 Length: 578
 Ratio: 3.440 Gaps: 15
 Percent Similarity: 68.166 Percent Identity: 48.962

alignment_block:

US-09-171-553b-4 x US-09-011-745-2

Align seq 1/1 to: US-09-011-745-2 from: 1 to: 7616

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
 728 ATGGCCAGACTGTTACCACTCCCTTAAGTTGACCTTAGCTACCTGGAA 777
 17 GlnuValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGly 34
 778 AGATGTGAGCGGAGATGCTACACACAGATCGTATGATCAAGAGAGAC 827
 34 rofTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 828 GTTGGGTTCCTTCTGCTCTGCTGCAATGCGCAACCTTAACTGATGAG 877
 51 ProSerGluGlyThrPheAsnSerGluIleIleLeuAlaValLysAlaI 67
 878 CCGCAGAGCGGACCTTTAACCGAGACCTCATCACAGGTTAGAGACAA 927
 67 eIlePheGlnThrGlyProGlySerHisProAspGlnLysProTyrIle 84
 928 GGTCTTTTACCTGCGCGCATGAGACACAGACAGAGTCCCTCATCATCG 977
 84 euThrTrpGlnAspLeuAlaGluAspProProProTyrLysProTrp 100
 978 TGACCTGGAGAGCTTGCTGTTTACCCCTCTGCTGCTGCTGCTGCTGCT 1027
 101 LeuAsnLysProAlaGlyProGlyProArgIleLeuAlaLeuGluLys 117
 1028 GTACAC...CCTAAGCCTCCGCTCCCT 1051
 117 sasnLysHisSerAlaGluLysValGluProSerSerTyrLeuPro 134
 1052CTTCCTCATCTCCGCTCCCTCCCTCC 1079
 134 rgaAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
 1080 TTGAACCTCTCTGTTGACCCGCTCGATCTCCCTTATATCACCCCTC 1129
 144 ThrProThrCysSerProThrPro.....LeuSerSerThrG1 156
 1130 ACTCTTCTCTAGGCGCAACCTTAACCTCAAGTTCTTTCGACAGTGG 1179
 156 yCysCysGluGly.....ThrSerAlaProPro.... 165
 1180 G.....GGGCGGCTCATGACCTACTTACAGAAACCCCGGCTT 1220
 165 165
 1221 ATAGGAGCCCAAGACACCCCTTCCGACAGGAGCAATGCTGGAGAA 1270
 166 GlyAlaProValValGluGlyPro.....AlaAlaGlyThr 177
 1271 GGGACCCCTGCGGAGAGAGACCGACCCCTCCCAATGACATCTCGCT 1320
 177 rArgSerArgArgLysAlaThrProGluArgThrAspGluIleAlaIle 194
 1321 AGGTGGAGAGCGGAGAGCCCTGTGGCGGACCTCCTACTCTGCGAGCAT 1370

```

194 eupProLeuAArgThrTyrlGlyProProMetProGlyGlnLeuGlnPro 210
   ::::::::::::::::::::
1371 TCCCCCTCCCGCAGAGAGA.....AACGACAG..... 1399
211 LeuGlnTyrlTrpProPheSerSerAlaAspLeuTyrlAsnTrpLysThrAs 227
   ::::::::::::::::::::
1400 CTTCATACATGCGCGCTTCCTCTTCGTGACCTTACCAACGAGAAATAA 1449
227 nHisProProPheSerGlnAspProGlnArgLeuThrGlyLeuValGln 244
   ::::::::::::::::::::
1450 TAAACCTCTCTTTCTGAGATCCAGGTAAACTACAGCTGTAGTCAGT 1499
244 erLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
   ::::::::::::::::::::
1500 CTGTTCATCATCCCATCAGCCACCTGGCAGCAGCTGTACAGACCTGTG 1549
261 GlnThrLeuPheThrThrGlnGlnArgGlnArgGlnLeuGlnValAar 277
   ::::::::::::::::::::
1550 GGGACTCTGCTGACCGGAGAGAAACAAACGGGTGCTGTAGAGGCTAG 1599
277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGln 294
   ::::::::::::::::::::
1600 AAAGCGGTGGCGGCGCATGATGGCGGCCCACTCACTGCCAATGAG 1649
294 leAspMetGlyPheProLeuThrArgProGlyTrpAspTrpAsnThrAla 310
   ::::::::::::::::::::
1650 TCGATGCCGCTTTCCTCCGAGCGCCAGACTGGGATTACACACCCAG 1699
311 GlnGlyArgGlnSerLeuLysIleThrArgGlnAlaLeuValAlaGlyLe 327
   ::::::::::::::::::::
1700 GCAGGTAGAGAACCACTAGTCCACTATGCCAGTTCCTCTCCTACCGGCT 1749
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGlnValM 344
   ::::::::::::::::::::
1750 CCAAAACGGGGGAGAGAGCCCAATTTGGCCAAAGGTAAAGAAATAA 1799
344 etGlnGlyProAsnGlnProProSerValPheLeuGlnArgLeuMetIu 360
   ::::::::::::::::::::
1800 CACAAGGGGCCAATGAGTCTCCCTGCGCTCTCTGAGAGACTTAAGGAA 1849
361 AlaPheArgArgPheThrProPheAspProThrSerGlnAlaGlnLysAl 377
   ::::::::::::::::::::
1850 GCCTATCGGAGGTACACTCTTATGACCTGAGAGCCAGGCGCAGAAAC 1899
377 aserValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
   ::::::::::::::::::::
1900 TAATGTCTATGCTTTCATTGTCAGTCCCGACATTTGGGAGAA 1949
394 yLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGlnLeuArgAspLeuVal 410
   ::::::::::::::::::::
1950 AGTTAGAGAGGTTAGAGATTAAACAAACAGCGCTTGAGATTGGTT 1999
411 ArgGlnAlaGlnLysValTyrlTrpArgArgGlnThrGlnGlnLysGln 427
   ::::::::::::::::::::
2000 AGAGAGGACGAAAGATCTTTAATAACGAAACCCCGAGAGAAAGAGA 2049
427 uGlnArgLysGlnLysGlnArgGlnArgGlnArgGlnArgGlnArgGln 440
   ::::::::::::::::::::
2050 GGAAGCTATCTAGGAGAGAAACAGAGAAAGAGAGAGCGCTGAGACAG 2099
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
   ::::::::::::::::::::
2100 AGGATGAGCGAAGAAAGAAAGAAAGATGCTGAGAGACTAGAGAGATG 2149
450 ThrLysIleLeuAlaAlaValAlaGlnGlyLysSerSerArgGlnArgGln 466
   ::::::::::::::::::::
2150 AGCAAGCTATTTGGCAGCTGCTGTAGTGACAGAAACAG..... 2188
466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeu 483
   ::::::::::::::::::::
2189 .....GATAGCAGAGGAGAGAA...C 2207
483 LysAsnArgThrProLeuAspLysAspGlnCysAlaTyrlCysLysGlnLys 499

```

```

2208 GAAAGAGGTCCCACTCGATTCGAGACAGTGTGCTACTGCAAGAAAG 2257
   ::::::::::::::::::::
500 GlyHisTrpAlaArgAsnCysProLysLys.....GlyAsnLysGlyPr 514
   ::::::::::::::::::::
2258 GGGCACTGGGCTAAAGATTGTCCCAAGAAACAGAGAGACCTCGGGACC 2307
514 oLys.....ValLeuAlaLeuGlnGln 521
   ::::::::::::::::::::
2308 AAGACCCAGACCTCCCTCTGACCTTAGATGAC 2341
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-850-961-1

```

seq_documentation_block:

```

/ Sequence 1, Application US/08850961
/ Patent No. 6013517

```

GENERAL INFORMATION:

```

/ APPLICANT: Respass, James G.
/ APPLICANT: De Polo, Nicholas J.
/ APPLICANT: Chada, Sunil
/ APPLICANT: Sauter, Sybille
/ APPLICANT: Bodner, Mordechai
/ APPLICANT: Driver, David A.
/ TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
/ NUMBER OF SEQUENCES: 45
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Chiron Corporation, Intellectual Property - R440
/ STREET: P.O. Box 8097
/ CITY: Emeryville
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94662-8097

```

COMPUTER READABLE FORM:

```

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/850,961
/ FILING DATE: 05-MAY-1997
/ CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:

```

/ NAME: Kruse, No. 6013517/man J.
/ REGISTRATION NUMBER: 35,235
/ REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (510) 601-3520
/ TELEFAX: (510) 655-3542
/ INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:

```

/ LENGTH: 8332 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-850-961-1

```

alignment_scores:

```

Quality: 1355.50 Length: 578
Ratio: 3.440 Gaps: 15
Percent Similarity: 68.166 Percent Identity: 48.962

```

alignment_block:

```

US-09-171-553b-4 x US-08-850-961-1 ..

```

```

Align seg 1/1 to: US-08-850-961-1 from: 1 to: 8332

```

```

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpTh 17
   ::::::::::::::::::::
621 ATGGCGGAGACTGTATACCACTCCCTTAAGTTGACCTTAGCTACATGGAA 670
17 rGlnuValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGly 34
   ::::::::::::::::::::

```

```

671 AGATGTCGAGCGATCGCTCAACACCATCGGTAGATGTCAAGAGAC 720
34  rdtgltglnthrphecyasalaserglutrprothrphaspvalglttrp 50
721 GTTGGGTTACCTTTCGCTCTGAGATGCGCAACCTTTAAGCTGAGTGG 770
51  ProsergltglutThrPheasnserglutlleleuAlaValLysAlaIle 67
771 CCGGACGAGCGACCTTTAACCGAGACCTCATCACCGAGTTAAGATCAA 820
67  eilephecglutThrGlySerHisproaspglutluProtyrllel 84
821 GGTCTTTTACCTGGCCGATGAGACACCGACAGAGTCCCTACATCG 870
84  eutThrGlnAspleuAlaGluAspProProProProValLysProTrrp 100
871 TGACCTGGGAAGCCTTGCTTTTGAACCCCTCCCTGGTCAAGCCCTTT 920
101  LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyLys 117
921 GTACAC...CCTAAGCCTCGGCTCT... 944
117  sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrrLeuProA 134
945 .....CTTCCTCATCCGCCCGCTCTCCGCC 972
134  rGAsp.....ArgGlyAlaAlaAspleuAlaGly 143
973 TTGAACCTCCTCGTTCGACCCCGCTCGATCCCTTTATCAGCCCTC 1022
144  ThrProThrCysSerProThrPro.....LeuSerSerThrG1 156
1023 ACTCCTCTCTAGCGCCAACTTAACCTCAAGTCTTTCGACAGTGG 1072
156  yCysCysGluGly.....ThSerAlaProPro... 165
1073 G.....GGGCCCTCATCGACCTACTTACAGAAAGCCCGCCTT 1113
165 ..... 165
1114 ATAGGACCCCAAGACACCCCTCCGACAGAGGAGAAAGTGAGAA 1163
166  GlyAlaProValAlaGluGlyPro.....AlaAlaGlyTh 177
1164 GGCACCCCTCGGAGAGGACCGACCCCTCCCAATGGCATCTCGCT 1213
177  rArgSerArgArgGlyAlaThrProGluArgThrAspGluIleAlaIle 194
1214 AGGTGGAGACGAGAGCCCTGTGGCCGACTCCACTCTCGACAGCAT 1263
194  euProLeuArgThrTyrrGlyProPrometProGlyGlnLeuGlnPro 210
1264 TCCCTCCGCGCAGAGAGA.....AAGGACAG..... 1292
211  LeuGlnTyrrProPheSerSerAlaAspleuTyrrAsnTrpLysTras 227
1293 CTTTCAATATGCGCTTCTCTCTGACCTTTAACAACCTGGAATAATA 1342
227  nhIsProProPheSerGluAspProGlnArgLeuThrGlyLeuValGlu 244
1343 TAAACCTCTTTTGTGAAGATCCAGGTAACGTACAGACTGTGATCAGT 1392
244  eLLeuMetPheSerHisGlnProThrTrpAspCysGlnGlnLeuLeu 260
1393 CTGTTCTCATCAACCATCAGCCACCTGGGAGCATCTCAGCAGCTTGG 1442
261  GlnThrLeuPheThrThrGluGluArgGluGlyLeuLeuGluAlaAr 277
1443 GGGACTCTGCTGACCGAGAGAAACAACGGTCTCTTAAAGAGCTAG 1492
277  gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGlu 294
1493 AAAGGCGGTGGGGGATGATGGGGCCCACTCAACTGCCCCAATGAAG 1542

```

```

294  leAspMetGlyPheProLeuThrArgProGlyTrpAspTyrrAsnThrAla 310
1543 TCGATGCCGCTTTCCTCCGAGGCCCGAGCTGATGATTAACACCCAG 1592
311  GluGlyArgGluSerLeuLysIleTyrrArgGlnAlaLeuValAlaGlyLe 327
1593 GCAGGTAGAACCACTAGTCCACTATCCGCAAGTGTCTCTACGGGCT 1642
327  uArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValM 344
1643 CCAAAACCGCGGAGAGCCCAACCAATTTGCCAAGCTAAAGAAATA 1692
344  eGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
1693 CACAAGGGCCCAATGACTCTCTCGGCTTCTTCAAGACACTTAAGAA 1742
361  AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAl 377
1743 GCCTATCGCAGGTACACTCTTATGACCTGAGACCCAGGCGAAGAAC 1792
377  aSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysL 394
1793 TAATGTCTGATGCTTTCATTGGCAGTCTGCCAGACATTGGAGAA 1842
394  yLsLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeuArgAspleuVal 410
1843 AGTTAGAGAGAGTTAGAAATTAACAAACAGACGCTTGAGATTTGGT 1892
411  ArgGluAlaGluLysValTyrrArgArgGluThrGluGluGlyG1 427
1893 AGAGAGCAAAAGATCTTAAATAACGAAACCCCGAGAAAGAAAGA 1942
427  uGlnArgLysGluLysGluArgGluGluArgGluGluArg... 440
1943 GGAAGCTATCAGAGAGAAACAGAGAAAGAAAGAACCGCTAGACAG 1992
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
1993 AGGATGACGAGAAAGAAAGAAAGATCTAGAGACATRAGACATG 2042
450  ThrLysIleLeuAlaAlaValAlaGluLysSerSerArgGluArgG1 466
2043 AGCAACCTATTGGCCACTGCTGTAGTACAGAAACAG..... 2081
466  uArgAspPheArgLysIleArgSerGlyProArgGlnSerIleLysLeuG 483
2082 .....GATAGACAGGAGAGAA...C 2100
483  LysAsnArgThrProLeuAspLysAspGlnCysAlaTyrrCysLysGluLys 499
2101 GAAGAGAGTCCCACTCGATCGCGCAGTGTGCTCTACTCTCAAGAAAG 2150
500  GlyHISTPAlaArgAsnCysProLysLys.....GlyAsnLysGlyPr 514
2151 GGGCACTGGCTTAAAGATTGTCCCAAGAAACACGAGACCTGGGGAGC 2200
514  oLys.....ValLeuAlaLeuGluGlu 521
2201 AAGACCCGAGACCTCCCTCGACCTTAGATGAC 2234

seq_name: /cogn2_6/prodata/2/lna/5A_COMB.seq:US-08-258-420-13
seq_documentation_block:
; Sequence 13, Application US/08258420
; Patent No. 5710037
; GENERAL INFORMATION:
; APPLICANT: Menhais, Arthur W.
; TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Ve
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gillillian, Cecchi, Stewart & Olstein

```

```

? STREET: 6 Becker Farm Road
? CITY: Roseland
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 inch diskette
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WordPerfect 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/258,420
? FILING DATE: 10-JUN-1994
? CLASSIFICATION: 435
? PRIORITY APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Olstein, Elliot M.
? REGISTRATION NUMBER: 24,025
? REFERENCE/DOCKET NUMBER: 271010-208
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8202 bases
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: genomic DNA
? FEATURE:
? NAME/KEY: viral genome
? US-08-258-420-13

```

```

alignment_scores:
  Quality: 1354.50      Length: 578
  Ratio: 3.438          Gaps: 15
  Percent Similarity: 68.166  Percent Identity: 48.789

```

alignment_block:

US-09-171-553B-4 x US-08-258-420-13/rev ..

Align seg 1/1 to reverse of: US-08-258-420-13 from: 1 to: 8202

```

1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisThrPth 17
  |||||
7146 ATGGCCGAGCTGTTACACCTCCTTAAGTTGACCTTAGCTCAGTGA 7097
  |||||
17 GlnValArgSerArgAlaHisAsnLeuSerValGlnValLysLysGlyP 34
  |||||
7096 AGATGTCAGCGGATCGCTCACACCACTCGTATGATGTCAGAGAGAC 7047
  |||||
34 rortGlnThrPheCysAlaSerGlnThrProThrPheAspValGlyTrp 50
  |||||
7046 GTTGGGTACTCTGCTGTCAGAAATGGCCACCTTAACATGATGATG 6997
  |||||
51 ProSerGlnGlyThrPheAsnSerGlnLeuLeuLeuAlaValLysAla 67
  |||||
6996 CCGGAGACGGCACCCTTTAAACCGAGACCTCATCACCCAGGTTAAGATCA 6947
  |||||
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  |||||
6946 GGTCTTTTACCTGGCCCGCATGAGACCCAGACAGGTCCTCATCATCG 6897
  |||||
84 eurtThrGlnAspLeuAlaGlnAspProProProThrValLysProTrp 100
  |||||
6896 TGACCTGGAGAGCTTGGCTTTTACCCCTCCCTGGTCAAGACCTTT 6847
  |||||
101 LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGly 117
  |||||
6846 GTACAC...CCTAAGCTTCGCTCTCT..... 6823

```

```

117 sAsnLysHisSerAlaGlnLysValGlnProSerSerSeryrLeuProA 134
  |||||
6822 .....CTTCCTCCATCCGCGCCGCTCTCTCC 6795
  |||||
134 rGAsp.....ArgGlyAlaAlaAspLeuAlaGly 143
  |||||
6794 TTGAACCTCTCGTTCGATGACCCCGCTCTCTCTTATTCACACCTCC 6745
  |||||
144 ThrProThrCysSerProThrPro.....LeuSerSerThrGln 156
  |||||
6744 ACCTCTCTCTAGGCGCCAAACCTCAAGTTCTTTCGACAGTGG 6695
  |||||
156 yCysCysGlnGly.....ThrsAlaProPro... 165
  |||||
6694 G.....GGGCGCTCATGACCTACTTACAGAGACCCCGCTT 6654
  |||||
165 ..... 165
6653 ATAGGACCCCAAGACCCCTTCGACAGGAGGAGGAATGATGAGAA 6604
  |||||
166 GlyAlaProValValGlnGlyPro.....AlaAlaGlyThr 177
  |||||
6603 GCGACCCCTCGGAGAGGACCGACCCCTCCCAATGGCATCTCGCT 6554
  |||||
177 rArgSerArgArgGlyAlaThrProGlnLysArgGlnAlaLeu 194
  |||||
6553 ACCTGGGAGACGGGACCCCTGTGGCGACATCTACTCTGCGAGCAT 6504
  |||||
194 eudProLeuArgThrTyrgLysProPrometProGlyGlnLeuGlnPro 210
  |||||
6503 TCCCTCCGCGCAGGAGGA.....AACGGACG..... 6475
  |||||
211 LeuGlnLysThrProPheSerSerAlaAspLeuTyraAsnTrpLysThr 227
  |||||
6474 CTCAATACCTGCGGCTCTCTCTCTGACCTTTACACTGGAATAATA 6425
  |||||
227 nHisProProPheSerGlnAspProGlnArgLeuThrGlnLeuValGln 244
  |||||
6424 TAACCTCTCTTTTCTGAAGATCCAGTAAACGACGCTCATGATGACT 6375
  |||||
244 erLeuMetPheSerHisGlnProThrTrpAspPcysGlnGlnLeuLeu 260
  |||||
6374 CTGCTTCATACACCATCCAGCCACCTGGGAGACATGTCAGACGCTTG 6325
  |||||
261 GlnThrLeuPheThrThrGlnGlnArgGlnArgGlnLeuGlnAla 277
  |||||
6324 GGGACTCTGTGACCGGAGAGAAACAAACGAGGTCTCTTGAAGCTAG 6275
  |||||
277 gLysAsnValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGln 294
  |||||
6274 AAGGGGCTGCGGGGAGATGATGGGCGCCCACTCAACGCCAATGAAG 6225
  |||||
294 leaSpMetGlyPheProLeuThrArgProGlyTrpAspTyrAsnTrpAla 310
  |||||
6224 TCGATCGCGGTTTTCCTCTGACCGCCAGACGAGGATTACACACCCAG 6175
  |||||
311 GlnGlyArgGlnSerLeuLysLysLeuArgGlnAlaLeuValAlaGly 327
  |||||
6174 GCGAGTGAAGACCACTAGTCCACTATTCGCACTTCTCTTACGGGCTCT 6125
  |||||
327 uArgGlyAlaSerArgArgProThrAsnLeuAlaValArgGlnValM 344
  |||||
6124 CCAAAACGGGCGGAGAAAGCCCAATTTGGCAGAGTAAGAAAGATTA 6075
  |||||
344 etGlnGlyProAsnGlnProProSerValPheLeuGlnArgLeuMetGln 360
  |||||
6074 CACAAAGGGCCCAATGAGTCCCTCGCCCTCTCAAGAGACTTAAGAA 6025
  |||||
361 AlaPheArgArgPheThrProPheAspProThrSerGlnAlaGlnLys 377
  |||||
6024 GCCTATCGCAGTACCTCTTATGACCTGAGGACCCAGGCGCAAGAAC 5975

```

```

377 aserValAlaLeuAlaPheIleGIyGlnSerAlaLeuAspIleArgLysL 394
5974 TAATGCTGATATGCTTTTCAATTTGGCAGCTCCCGACATTTGGAGAA 5925
394 yslLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGlnLeuArgAspLeuVal 410
5924 AGTTAGAGAGGTTAGAGATTAAATAACAACAGCCCTTGACATTTGGTT 5875
411 ARGGLuAlaGlnLysValTyrTyrArgArgGlnThrGlnGlnLysGln 427
5874 AGAGAGGCGAGAAAGATCTTTAATAACGAGAAACCCCGGAGAAAGAGA 5825
427 uGlnArgLysGlnLysGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 440
5824 GGAACCTATCAGAGAGAGAAACAGAGATCGTAGAGAGACATAGAGAGATG 5775
441 .....ArgAspArgArgGlnGlnLysAsnLeu 449
5774 AGGATGAGCAGAAAGAAAGAAAGATCGTAGAGAGACATAGAGAGATG 5725
450 ThrLysIleLeuAlaAlaValAlaGlnLysSerSerArgGlnArgGln 466
5724 AGCAAGCTATTGGCCACTGCTGCTAGTGACAGAAACAG ..... 5686
466 uArgAspPheArgLysIleArgSerGlyProArgGlnSerGlnLysLeuG 483
5685 .....GATAGACAGGAGAGAGAA...C 5667
483 IYAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGlnLys 499
5666 GAGAGAGGTCCCACTCGATCCGACGACGATGCTGCTACTGCAAAAGAAAG 5617
500 GlyHisThrPalaArgAsnCysProLysLys.....GlyAsnLysGlyPr 514
5616 GGGGCACTGGGCTAAGATTGTCCCAAGAAACACAGAGAGACCTCGGGGAC 5567
514 OLYS.....ValLeuAlaLeuGlnGln 521
5566 AAGACCCAGACACTCTCTGACCTGACCTAGATGAC 5533
seq_name: /cgn2_6/prodata/2/1na/3A_COMB.seq:US-08-110-300A-8
seq documentation block:
; Sequence 8, Application US/08110300A
; Patent No. 5643736
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,300A
; FILING DATE: 20-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461

```

```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-110-300A-8

alignment_scores:
    quality: 1350.50      length: 589
    ratio: 3.535          gaps: 14
    Percent Similarity: 64.856      Percent Identity: 47.538

alignment_block:
US-09-171-553B-4 x US-08-110-300A-8 ..

Align seq 1/1 to: US-08-110-300A-8 from: 1 to: 8323

1 MetGlyLinhThrValThrThrProLeuSerLeuThrLeuAspHisThrph 17
|||||.....
619 ATGGGCGAGGCTGTACACCCCTTAAGTTGACTTAAACACCTGAGAA 668
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
|||||.....
669 GGATGTGAACGGACAGACCCACAACTGCTGTAGAGGTTAGAAAAGC 718
34 rOTrGlnThrPheCysAlaSerGlnTyrProThrPheAspValGlyTyr 50
|||||.....
719 GCTGGGTTCATTCTGCTGTGAGAAAGGCCAACCTTCAACGTCGATGG 768
51 ProSerGlnLysThrPheAsnSerGlnIleIleLeuAlaValLysAla 67
|||||.....
769 CCAAGAGCGGACCTTTTAAACCCAGATTATTACAGAGTTAAGATCAA 818
67 eIlePheGlnThrGlyProGlySerHisProAspGlnLuproTyrIle 84
|||||.....
819 GGTCTTCTACCTCGGCCACATGACATCGGATCGATCGATCCCTACATCG 868
84 eutThrTrpGlnAspLeuAlaGlnAspProProTyrValLysProTyr 100
|||||.....
869 TGACCTGGGAAGCTATAGCAGTAGACCCCTCGGTGGACAGACCTTC 918
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlnLys 117
|||||.....
919 GTGCAC...CTTAACCTCC.....CTCTCTCT ..... 945
117 sAsnLysHisSerAlaGlnLysValGlnProSerSerSerTyrLeuPro 134
|||||.....
946 .....CCCCCTACAGCCCTCTCTCCAC 970
134 rGAspArgGlnAlaAlaAspLeuAlaGlnThrProThrCysSerProThr 150
|||||.....
971 CTGAA.....CCCCACTCTGACCCCG 993
151 ProLeuSerSerThrLysCysGlnGlnLysThrSerAlaProProGlnAl 167
|||||.....
994 CCCCAGTCTCTCCCTC.....TA 1010
167 aProValAlaGlnLysProAlaAlaGlnThrArgSerArgGlnAlaT 184
|||||.....
1011 TCCGGCTCTCACTTCTCT...TTAAACCAACCAACTAGGCTCAAGTCC 1057
184 hProGlnArgThrAspGlnIleAlaIleLeu..... 194
|||||.....
1058 TTCCTGATAGCGAGAGACCATCATTTAGTCTACGAGAGACCTCGG 1107
195 ProLeuArgThrTyrGlyProProMetProGlyGly..... 206
|||||.....
1108 CCTTACCGGGACCCAGGGCCACTCTCTCTGACGGGAACGGCATAGCGG 1157

```

```

206 ..... 206
1158 AGAAGTGGCCCTACAGAGAGCCCTGACCTTCCCAATGATATCCC 1207
207 .....Gln 207
1208 GCCTGGCGGGAGAAAGAACCCCTGGCGGATTTACTACTACTCTCAG 1257
208 LeuGlnProLeu.....GlnTyrTrpProh 216
1258 GCGTTCCTCCCTTCGCTGGAGGAGATGACATATCATATCGCCATT 1307
216 eSerSerAlaAspLeuTyrAsnTyrPylsThrAsnHisProProhSerG 233
1308 TTCCTCTCTGACCTCTAATCGAATAAATACACACCCCTCTTCCG 1357
233 LuAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHis 249
1358 AGGACCCACCTAATTGACAGCTTGATCGAGTCGCTCTCTACTCAT 1407
250 GlnProThrTTPAspAspCysGlnGlnLeuLeuGlnThrLeuPheThr 266
1408 CAGCCCACTTGGATGACACGCAACAGCTATTAGGAGCCCTGCTGACGG 1457
266 rGluGluArgGluArgIleLeuLeuGlnAlaArgLysAsnValProGly 283
1458 AGAAGAAACACAGCGAGTCTCTAGAGCCGCAAGGCGGTTGAGGGG 1507
283 LaAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMetGlyPhePro 299
1508 AGACGAGAGCCCACTGAGTCCCAAGACATTAATGATGCTTTCC 1557
300 LeuThrArgProGlyTyrPylsThrAlaGluArgLysSerLe 316
1558 TTGGAGCTCCGACGCTGGACTCAACACCAAGAGTAGAGACACACT 1607
316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
1608 AGTCCACTATTCGACAGTTGCTCCTAGCGGTCTCCAAACGCGGAGAA 1657
333 rGpProThrAsnLeuAlaLysValArgLysValMetGlnGlyProAsnGlu 349
1658 GCCCACCACATTGGCCAGAGTAAGAGGATTAACCCAGGAGCTATGAG 1707
350 ProProSerValPheLeuGlnArgLeuMetGluAlaPheArgArgPheTh 366
1708 TCTCCCTCAGCGCTTTTAGAGAGACTCAAGAGGCTATCGCAGATAC 1757
366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaP 383
1758 TCCTTATGACCTGAGGAGCCAGGGCAAGAAACCAATAGTGGCATGTGCAT 1807
383 heIleGlyLysSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
1808 TCATGTGCACTCCGCCCGCGGATATCGGGCAAAAGCTTAGAGCGTTGAA 1857
400 GlyLeuGlnGluAlaGlnLeuArgAspLeuValArgGlnAlaGluLysVa 416
1858 GATTGAGAGAGTAAGACCTTAGGAGACTTAAGAGGAGAGCTGAAAGAT 1907
416 lTyrTyrArgArgGluThrGlnGluGluLysGlnArgLysLysG 433
1908 CTTTAAATAACGAGAAACCCCGGAAAGAAAGAGAAAGCTATTAGAGAG 1957
433 LuArgGluGluArgGluArg..... 440
1958 AAACAGAGGAAAGAAAGAACCCCTAGGCGAGAGATGTCCAGAGAGAG 2007
441 .....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAl 455
2008 AAGGAGAGGAGCGCAGAAACATAGAGAAATGACTTAAGTTGCTGCGCTAC 2057

```

```

455 aValValGluGlyLysSerSerArgGluArgGluArgAspPheArgLysI 472
2058 TGCTGTTAGCGCGGAC.....AGACAGG 2080
472 lArgSerGlyProArgGlnSerGlyAsnLeuGlnLysAsnArgThrProLeu 488
2081 ATAGACAGGAGAGAGAGCA.....AGAGGCCCAACATC 2115
489 AspLysAspGlnCysAlaTyrCysLysGluLysGlyHisThrAlaArgAs 505
2116 GACCACGACAGCTGTGCTACTGCAAGAAAGGACATTTGGCTAGAGA 2165
505 nCysProLys.....GlyAsnLysGlyProLys.....V 516
2166 TTGCCCAAGAAAGCAAGAGACCCCGGAGACACAGCCCAAGCCCTGCC 2215
516 aLeuAlaLeuGlnGlu 521
2216 TCCTGACCTTAGACGAT 2232

seq_name: /cogn2_6/ptodata/2/ina/5R_COMB.seq:us-08-886-642-8

seq_documentation_block:
? Sequence 8, Application US/0886642
? Patent No. 5952474
? GENERAL INFORMATION:
? APPLICANT: Pinter, Abraham
? APPLICANT: Kayman, Samuel
? TITLE OF INVENTION: FUSION GLYCOPROTEINS
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson P.C.
? STREET: 45 Rockefeller Plaza, Suite 2800
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/886,642
? FILING DATE: 01-JUL-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/110,300
? FILING DATE: 20-AUG-1993
? APPLICATION NUMBER: 07/938,100
? FILING DATE: 28-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Hone, William J.
? REGISTRATION NUMBER: 26,739
? REFERENCE/DOCKET NUMBER: 07763/010002
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212/765-5070
? TELEFAX: 212/258-2291
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8323 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? US-08-886-642-8

alignment_scores:
? Quality: 1350.50 Length: 589
? Ratio: 3.535 Gaps: 14
? Percent Similarity: 64.856 Percent Identity: 47.538

alignment_block:

```


US-09-171-553B-4 x US-08-886-642-8 ..

Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

```

1 MetGlyGlnThrValThrProLeuSerLeuThrLeuSphistPth 17
  |||||
619 ATGGCCGAGGCTTACCCACCCCTTAAGTTGACTTAGACCACTGGAA 668
17 rgluValArgSerAlaHisAsnLeuSerValGlnValLysLysGly 34
  |||||
669 GGATGTGAGACGACGACCCCAACACCTGCGTAGAGATTAGAAAGGC 718
34 rorPGLnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
  |||||
719 GCTGGTTACATCTCTGCTGACAAATGGCCAACTTCAACGTCGATGG 768
51 ProSerGlnGlyThrPheAsnSerGlnIleLeuAlaValLysAlaI 67
  |||||
769 CCAGGACGAGCGACTTTAACCCAGACATTATACACAGATTAGATCAA 818
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
  |||||
819 GGTCTTCTACCTGGCCACATGACATCGGATCAGTCCCTCATCATCG 868
84 eutHrTrpGlnAspLeuAlaGluAspProProProProValLysProTrp 100
  |||||
869 TGACCTGGAGAGCTATAGCAGTAGACCCCTCTCCGCTGAGACCTTC 918
101 LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGly 117
  |||||
919 GTGCAC...CTTAACCTCC...CTCTCTCT... 945
117 sasnLysHisSerAlaGluLysValGlnProSerSerSerTyrLeuPro 134
  |||||
946 .....CCCCCTGACGCCCTCTCTCCAC 970
134 rgaAspArgLysAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
  |||||
971 CTGAA.....CCCCACTCTCCACCC 993
151 ProLeuSerSerThrGlyCysCysGlnGlyThrSerAlaProProGlyAl 167
  |||||
994 CCCCAGCTCTCCCTC.....TA 1010
167 aProValAlaGlnGlyProAlaAlaGlyThrArgSerArgArgLysAla 184
  |||||
1011 TCCGGCTCTCACTCTCTCT...TTAAACCAACCACTAGGCTCAAGTCC 1057
184 hrProGluArgThrAspLeuIleAlaIleLeu..... 194
  |||||
1058 TTCCTGATAGCGGAGGACCACTCATTTGATCTACAGGAGACCTCTCC 1107
195 ProLeuArgThrTyrGlyProPheProGlyGly..... 206
  |||||
1108 CTTTACCGGAGACCGGACCCCTCTCTGACGAGGAGACGCGATACGG 1157
206 ..... 206
1158 AGAAGTGGCCCTACAGAGAGAGCCCTGACCTCCCAATGTATGCC 1207
207 .....Gln 207
1208 GCTGTGGGGGAAAGAAAGACCCCGTGCGGATTCTACTACTCTTCG 1257
208 LeuGlnProLeu.....GlnTyrTrpProh 216
  |||||
1258 GCGTTCCTCCCTGCGGAGGAGATGACAGTATCACTACTGCGCATTT 1307
216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProPheSerG 233
  |||||
1308 TTCTCTCTCTGACCTCTATTAACGAAATAACAACCCCTTCTCTCCG 1357
233 lAspProGlnArgLeuThrGlyLeuValGlnSerLeuMetPheSerHis 249
  |||||
1358 AGGACCCAGTAAATTTGACAGCTTTGATCGAGTCCCTTCTCTACTCAT 1407
250 GlnProThrTrpPaspAspCysGlnGlnLeuGlnGlnThrLeuPheThr 266
  |||||
1408 CAGCCACTTGGGATGATGACCAACACTATTAGAGGACCTGCTGACGG 1457
266 rgluGluArgLysArgLysLeuGlnAlaArgLysAsnValProGlyA 283
  |||||
1458 AGAAGAAAAACAGGAGATGCTCTAGAGGCCCAAGAGGCGGTGAGGG 1507
283 laAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMetGlyPhePro 299
  |||||
1508 AGGACGAGACCCCAACTCAGCTCCCAATGATTAATGATGCTTTGCC 1557
300 LeuThrArgProGlyTrpAspTyrAsnThrAlaGlnGlyArgGlnSer 316
  |||||
1558 TTGGACGTCCTCCGACTGGACTGACCTACACACCAAGAGTAGAACCCCT 1607
316 ulysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
  |||||
1608 AGTCCATATCGCCAGTCTCTCTACGGGCTCCAAAGCGGGCAGAA 1657
333 rProThrAsnLeuAlaLysValArgLysAlaMetGlnGlyProAsnGlu 349
  |||||
1658 GCCCACAACATTTGGCCAAAGTAAGATTAACCCAGGAGCTTAATGAG 1707
350 ProProSerValPheLeuGlnLysArgLeuMetGlnAlaPheArgArgPhe 366
  |||||
1708 TCTCCCTCAGCCTTTTAAAGAGACTCAAGAGGCTTTCAGATACAC 1757
366 rProPheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAla 383
  |||||
1758 TCCTTATGACCTGAGGACCCAGGCAAGAAACCAATGCGCATATCAT 1807
383 heIleGlyLysSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
  |||||
1808 TCATCTGCGAGTCGCCCGGATATCGGCGCAAGATTAGAGGCTTAGAA 1857
400 GlyLeuGlnGlnAlaGlnLeuArgAspLeuValArgGlnAlaGlnLysVa 416
  |||||
1858 GATTTGAAGAGTAAGACCTTAGAGACTTAGAGGAGCACTGCAAAAGAT 1907
416 lTyrTyrArgArgLysTrpGlnGlnGlnLysGlnArgLysGlnLysG 433
  |||||
1908 CTTTATTAAGAGAAACCCCGGAAAGAAAGAGAAAGATTTAGGAGAG 1957
433 lUArgGlnGlnLysArgLys..... 440
  |||||
1958 AAACAGAGGAAAGAAAGAACGCGTAGGCGACAGATGTGACAGAGAG 2007
441 .....ArgAspArgArgGlnGlnLysAsnLeuThrLysIleLeuAla 455
  |||||
2008 AAGGAGAGGAGACCGCAACAGATAGAAATAGTAAGTTGCTGCTAC 2057
455 aValValGlnGlyLysSerSerArgLysArgLysArgAspPheArgLys 472
  |||||
2058 TGTGTTAGGGGGGAG.....AGACAG 2080
472 leArgSerGlyProArgGlnSerGlyAsnLeuGlnLysAsnArgThrPro 488
  |||||
2081 ATAGACAGGAGAGAGAGCA.....AGGAGGCCCAACACTC 2115
489 AspLysAspGlnCysAlaTyrCysLysGlnLysGlnHisTrpAlaArg 505
  |||||
2116 GACCAGACAGCAGTGTCTACTAGCAAAAGGAGCATTTGGCTAGAGA 2165
505 nCysProLysLys.....GlyAsnLysGlyProLys.....V 516
  |||||
2166 TTCCCCCAAGAAAGCAAGAGAGACCCGGGACCAAGACCCAGGCTTCC 2215
516 alLeuAlaLeuGlnGlu 521
  |||||

```



```

300 LeuThrArgProGlyTyrPaspTyrAsnThrAlaGluGluArgGluSerLe 316
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1558 TTGGAGCTGCTCCGAGCTGAGCTACACACCCAGAGTAGAGACCACT 1607
316 ULysLeuTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1608 AGTCACATATGCGCAGTTCTCTCCAGGGGTCTCCAAACCGGGGAGAA 1657
333 rGProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsnGlu 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1658 GCCCCACCAATTTGGCCCAAGTAAAGAGTAACCCAGGACCTTAATGAG 1707
350 ProProSerValPheLeuGluArgLeuMetGlnAlaPheArgArgPheTh 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1708 TCCTCATACGCTTTTGTAGAGACTCAAGAGCGCTTATCGACATGATCAC 1757
366 rProPheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAlaP 383
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1758 TCCTTATACCTTGAGACCCAGGCGAAGAACCAATGTGGCCATGTGAT 1807
383 helLeuGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1808 TCATCTGGCAGTCCGCCCCGATATCCGGCGCAAGTTAGACCGCTTGAA 1857
400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGlyLysVa 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1858 GATTGGAAGCTAGACCTTGAGAGACTTACTGAGGGAGAGCTGAAAGAT 1907
416 LTyTyrArgArgGluThrGlnGluGluGluGluGluGluGluGluGlu 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1908 CTTTATATTAACGAGAACCCCGGAAAGAAAGAGAACGATTTAGGAGAG 1957
433 LuArgGluGluArgGluGluArg..... 440
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1958 AAACAGAGAAAGAGAACGCCGTAGGCGAGAGATGTCAGAGAGAG 2007
441 .....ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAla 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2008 AAGGAGAGGAGCGCGAAGACATAGAGAAATGAGTAAGTTGCGGTAC 2057
455 aValValGluGluLysSerSerArgGluArgGluArgAspPheArgLysI 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2058 TGCTGATGCGCGCAG.....AGACAGG 2080
472 leuArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProLeu 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2081 ATAGACAGAGAGAGAGCGCA.....AGGAGGCCCAACTC 2115
489 AspLysAspGlnGlyAlaTyrCysLysGluLysGlyHisThrPalaArgAs 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2116 GACCCAGCAGAGTGTGCTACTGCAAAAGAGGACATGTGGCTAGAGA 2165
505 nCysProLysLys.....GlyAsnLysGlyProLys.....V 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2166 TTGGCCCAAGAGCCAGAGAGACCCCGGAGACAGACCCAGGCGCTGCC 2215
516 alleuAlaLeuGluGlu 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2216 TCCTGACCTTAGACGAT 2232
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-110-300A-9
seq_documentation_block:
; Sequence 9, Application US/08110300A
; Patent No. 5643756
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSTON GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood
; STREET: 45 Rockefeller Pl.

```

```

? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/110,300A
? FILING DATE: 20-AUG-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Jacobs, Seth H.
? REGISTRATION NUMBER: 32,140
? REFERENCE/DOCKET NUMBER: 11698A50
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-757-2200
? TELEFAX: 212-586-1461
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10367 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? US-08-110-300A-9

alignment_scores:
    Quality: 1350.50      Length: 589
    Ratio: 3.535          Gaps: 14
    Percent Similarity: 64.856      Percent Identity: 47.538

alignment_block:
US-09-171-553b-4 x US-08-110-300A-9/rev ..

Align seg 1/1 to reverse of: US-08-110-300A-9 from: 1 to: 10367

1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisThr 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6977 ATGGGCGAGGCGTGTACCAACCCCTTAAGTTGACTTAGACCACTGAA 6928
17 rGluValArgSerArgAlaHisAsnLeuSerValGlnValLysGlyP 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6927 GGAATGCGAAGGAGACAGCCACACACTGTGCGTAGAGTTAGAAAAGGC 6878
34 rotPrpGlnThrPheGlyAlaSerGluThrProThrPheAspValGlyTrp 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6877 GCTGGGTATATCTCTCTGCGAAGTGGCCAACTTCAACGTGCGATGG 6828
51 ProSerGlnGlyThrPheAsnSerGluIleIleLeuAlaValLysAla 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6827 CCACGAGAGCGCACTTTTAACCCAGCATTTATACACAGTTAAAGTCAA 6778
67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrLeu 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6777 GGTCTTCTACCTGGCCCAATGACATCGGATCGAGTCAAGTCCCATATG 6728
84 eutThrPrpGlnAspLeuAlaGluAspProProThrValLysProTrp 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6727 TGACCTGGGAAGCTATAGCAGTAGACCCCTCCCTGGGCTAGACCTTC 6678
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6677 GTGCAC...CTTAACCTCC.....CTCTCTCT..... 6651
117 sAsnLysHisSerAlaGluLysValGluProSerSerSerTyrLeuPro 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6650 .....CCCCCTTACGCCCCCTCTCTCCAC 6626

```

134 rgaspargylalaalaspheulaglythrprothrycysserprothr 150
6625 CTGAA.....CCCCACTCTCGACCCG 6603
151 Proleuserthrlygylcyscylgllythrseralaproproglal 167
6602 CCCAGTCTCCCTC.....TA 6586
167 aProvalalaglulproalalaglthrargserargylalal 184
6585 TCCGCTCTCACTCTCTCT...TTAAACACCAAACTAGGCTCAAGTCC 6539
184 hrprogluargthraspjluilalalaleu..... 194
6538 TTCTGTATAGCGAGACCACTATGATCTACTACGAGACCTCCG 6489
195 Proleuargthrlygylproprometproglly..... 206
6488 CCTTACGGGAGCCAGGGCCACTCTCTCTGACGGAGGCGATAGCG 6439
206 206
6438 AGAGTGGCCCTACAGAGAGCCCTGACCTTCCCAATGATATCC 6389
207Gln 207
6388 GCGTGGGGGAGAAAGAACCCCGTGGCGGATTCTACTACCTCTAG 6339
208 LeuGlnProleu.....GlnTyrTrpProph 216
6338 GCGTCCCTTCCCTGGGAGGATGAGACATATCAATACGCGCAT 6289
216 eSerSerAlaspLeuTyrAsnTrpLysThrAsnHisProProph 233
6288 TTCTCTCTGACCTCTATTAAGTGAATAACACACCTCTTCTCCG 6239
233 LuaspProGlnArgLeuThrGlnLeuValGlnSerLeuMetPheSerHis 249
6238 AGGACCGACCTAAATGACACGCTTGTATGACGCTTCTCTTACTCAT 6189
250 GlnProThrTrpAspPcysGlnGlnLeuLeuGlnThrLeuPheThr 266
6188 CAGCCCACTGGATGATGACTGCCACACCTATTAGGACCCGCTGACGG 6139
266 rGlnGluArgGluArgGlnLeuLeuGlnValArgLysAsnValProGly 283
6138 AGAAGAAAAACAGCAGTGTCTCTAGAGCCGCAAGCGGCTCGAGGG 6089
283 LaaspGlyArgProThrGlnLeuGlnAsnGlnLeuLeuMetGlyPhePro 299
6088 AGGACGAGCCCACTGACTGCCCATGATGATTAATGATGCTTTTCCC 6039
300 LeuThrArgProGlyTyrPaspTyrAsnThrAlaGlnLysArgLysSer 316
6038 TTGGAAGCTCCGACTGGACTACACCAACGAGGTAGAGACCACT 5989
316 uLysLeuTyrArgGlnAlaLeuValAlaGlyLeuArgLysAlaSerArg 333
5988 AGTCACATATCGCAGTGTCTCTAGCGGCTCTCAAAACGCGGGAGAA 5939
333 rGProThrAsnLeuAlaLysValArgLysValMetGlnGlyProAsnGlu 349
5938 GCCCACCACATTTGGCCCAAGTAAAGGATTAACCCAGGAGCCTAATAG 5889
350 ProProSerValPheLeuGluArgLysLeuGlnAlaPheArgLysPhe 366
5888 TCTCCCTCAGCTTTTGTAGAGACTCAAGAGGCTATCCACATATGAT 5839
366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuLap 383
5838 TCTTATATACCTGTAGACCCAGGGCAAGAAACATGTGCGCATGTAT 5789

383 heIIeGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
5788 TCATCTGGCAGTCCGCCCGGATATCGGCGAAAGTTAGCGGCTTAGAA 5739
400 GlysLeuGlnAlaGlnLeuArgAspLeuValArgGluAlaGlnLys 416
5738 GATTGGAAGATAGACCTTAGAGACTTAGAGAGGAGGAGGAAAGAT 5689
416 lTyrTyrArgArgGluThrGlnGlnLysGlnGlnArgLysGlnLys 433
5688 CTTTAATTAACGAGAAACCCCGAAGAAAGAGAGACGATTATGAGAG 5639
433 LuArgGlnGluArgGlnGlnLysArg..... 440
5638 AAACAGAGAAAGAGAGAACCCGTAAGGCGAGAGATGTGAGAGAG 5589
441ArgAspArgArgGlnGlnLysAsnLeuThrLysIleLeuAla 455
5588 AAGAGAGGAGCGCCGAGAGACATAGAGAAATGATTAAGTCTGCTAC 5539
455 aValValGlnGlnLysSerSerArgGlnArgGlnArgLysPheArgLys 472
5538 TGTCTGTAGCGGCGAG.....AGCAGG 5516
472 leArgserGlyProArgGlnSerGlyAsnLeuGlyAsnArgThrProleu 488
5515 ATAGACAGGAGAGAGAGCA.....AGAGGCCCACTC 5481
489 AspLysAspGlnGlnLysAlaTyrCysLysGlnLysGlnLysTrpAlaArg 505
5480 GACACAGACAGCTGTCTCTACTGCAAGAAAGGACATGTTGGCTAGAGA 5431
505 ncysProLysLys.....GlyAsnLysGlnProLys.....V 516
5430 TTGCCCCAAGAACCAAGAGAGACCCCGGAGACAGACCCAGGCTTCC 5381
516 alLeuAlaLeuGlnGlu 521
5380 TCTGTACCTTAGACAT 5364
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-9
seq_documentation_block:
? Sequence 9, Application US/08886642
? Patent No. 5952474
? GENERAL INFORMATION:
? APPLICANT: Pinter, Abraham
? APPLICANT: Kayman, Samuel
? TITLE OF INVENTION: FUSION GLYCOPROTEINS
? NUMBER OF SEQUENCES: 25
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson P.C.
? STREET: 45 Rockefeller Plaza, Suite 2800
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows95
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/886,642
? FILING DATE: 01-JUL-1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/110,300
? FILING DATE: 20-AUG-1993
? APPLICATION NUMBER: 07/938,100
? FILING DATE: 28-AUG-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Hone, William J.
? REGISTRATION NUMBER: 26,739

REFERENCE/DOCKET NUMBER: 07763/010002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212/765-5070
 TELEFAX: 212/258-2291
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10367 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-08-886-642-9

alignment_scores:
 Quality: 1350.50 Length: 569
 Ratio: 3.535 Gaps: 14
 Percent Similarity: 64.856 Percent Identity: 47.538

alignment_block:

US-09-171-553b-4 x US-08-886-642-9/rev ..

Align seg 1/1 to reverse of: US-08-886-642-9 from: 1 to: 10367

1 MetGlglnThrValThrProLeuSerLeuThrLeuAspHisTrpTh 17
 |||||
 6977 ATGGGCGAGCGCTTACCCACCCCTTAAGTTGACTTGAACCACTGGA 6928
 17 rgluValArgSerArgAlaHisAsnLeuSerValGlnValLysGly 34
 |||||
 6927 GATGTCCAGACGACGACGCCACACCTCGTAGAGGTAGAAAAGC 6878
 34 rGTrGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTrp 50
 |||||
 6877 GGTGGGTACATCTCTCTCTGAGATGGCCACCTTCACGTGCGAGTG 6828
 51 ProSerGluGlyThrPheAsnSerGluLeuLeuValValLysAla 67
 |||||
 6827 CCACAGACGCGACTTTAACCCAGACATTATTACACAGGTAAAGATCA 6778
 67 eilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 84
 |||||
 6777 GGTCTCTCACCCTGGCCACCGACATCCGATCCGATCCCTCAATCG 6728
 84 eutHTrpGlnAspLeuAlaGluAspProProProTyrValLysProTyr 100
 |||||
 6727 TGACCTGGAGAGCTATAGACATAGACCCCTCCCTGGGTGACCCCTTC 6678
 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGly 117
 |||||
 6677 GTGCAC...CCTAAACCTCCC.....CTCTCTCT..... 6651
 117 sasnLysHisSerAlaGluLysValGluProSerSerTyrLeuProAla 134
 |||||
 6650CCCTTCAGCCCTCTCTCTCCAC 6626
 134 rgaSPArGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProThr 150
 |||||
 6625 CTGA.....CCCTTCAGCCCTCTCTCCAC 6603
 151 ProLeuSerSerThrGlyCysCysGluGlyThrSerAlaProProGlyAla 167
 |||||
 6602 CCCAGCTCTCCCTC.....TA 6586
 167 aProValValGluGlyProAlaAlaGlyThrArgSerArgArgValAla 184
 |||||
 6585 TCCGGCTCTCACTTCTCT...TTAAACAACAACCTAGCCCTCAAGTCC 6539
 184 hrProGluArgThrAspGluIleAlaIleLeu..... 194
 |||||
 6538 TTCTGTATGAGGAGACACCATATTGATCTACTACGAGAGACCTCCG 6489
 195 ProLeuArgThrTyrGlyProPrometProGlyGly..... 206

6488 CCTTACCGGACCCAGGCGCACCTCTCTACCGGAACGGCGATAGCG 6439
 206 206
 6438 AGAAGTGGCCCTACAGAGAGCCCTGACCTTCCCAATGGATGCC 6389
 207Gln 207
 6388 GCCTGGGGGAAAGAAAGAACCCCGGCGGATTTCTACTACCTCAG 6339
 208 LeuGlnProLeu.....GlnTyrTrpPro 216
 6338 GCGTTCCTCCCTTCCCTGGGAGGAAATGACAGATCAATACATGCGCAT 6289
 216 eSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSer 233
 6288 TTCTCTCTCTGACCTCTTAACTGGAATAATACACCCCTTTCTCCG 6239
 233 lAspProGlnArgLeuThrGlyLeuValGlnSerLeuMetPheSerHis 249
 6238 AGGACCCAGCTAAATTGACAGCTTATGAGTCCGTTCCTTACTCAT 6189
 250 GlnProThrTrpAspCysGlnGlnLeuLeuGlnThrLeuPheThrTh 266
 6188 CAGCCCACTGGGATGACTCCCAACACTATTATGAGACCTGCTGACGG 6139
 266 rGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGly 283
 6138 AGAAGAAACACAGAGACTGCTCTACAGGCGCCAAAGCGGTTGAGGG 6089
 283 lAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPhePro 299
 6088 AGAGCGAGCCCACTCAGCTCCCAATGACATTATGATGATTTTCC 6039
 300 LeuThrArgProGlyTyrPaspTyrAsnThrAlaGluGlyArgGlySer 316
 6038 TTGGAACTGCCAGCTGGGACTACACACCAACGAGTAGAGAACACCT 5989
 316 uLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
 5988 AGTCACATACCGCCACTGCTCTCTACCGGCTCCAAACGGGGAGAA 5939
 333 rProThrAsnLeuAlaLysValArgGluValMetGlnLysProAsnGlu 349
 5938 GCCCAACAATTTGGCCAAAGTAAGATACCCAGGACCTAATGAG 5889
 350 ProProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPhe 366
 5888 TCTCCCTCAGCTTTTAAAGAGACTCAAGGAGGCTTATCGCAGATAC 5839
 366 rProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAla 383
 5838 TCCTTATGACCTGAGAGACCCAGGCCAAGAACATATGCGCATGTAT 5789
 383 heIleGlyInSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGlu 399
 5788 TCATCTGCAATCGCCCGGATATCGGGCAAACTTGAAGCGGTAA 5739
 400 GlyLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLys 416
 5738 GATTGAAAGATTAAGACCTTAAAGAGCTTAAAGAGGAACTGAAAGAT 5689
 416 lTyrTyrArgArgGluThrGluGluLysGlnGlnArgLysGlyLysG 433
 5688 CTTTATTAACAGAGAAACCCGGAAGAAAGAGAGGATATTAGAGAG 5639
 433 lArgGluGluArgGluGluArg..... 440
 5638 AAACAGAGGAAAGAGAGACCGGTAGGCAAGATGTCCAGAGAG 5589
 441ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAla 455

```

5588 AAGGAGGGGACCGCAGACATAGAGAAATGAGTAAATTGCTGCTAC 5539
455 aValValGIuGIlySerSerArgGIuArgGIuArgAspHeArgLysI 472
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5538 TGTCGTTAGCGGGCAG.....AGACAGG 5516
472 leArgSerGIyProArgInSerGIyAsnLeuGIyAsnArgThrProLeu 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5515 ATAGACAGGGAGGAGAGCGA.....AGGAGGCCCACTC 5481
489 AspLysAspGIuGIyAlaTyrCysLysGIuLysGIyHisTPAlaArgAs 505
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5480 GACCAAGACAGTGTGCTACTGCAAGAAAGGACATGGGCTAGAGA 5431
505 nCysProLysLys.....GIyAsnLysGIyProLys.....V 516
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5430 TTGCCCCAAGAACCAAGAGACCCCGGGGACACGACCCCAAGCCTCC 5381
516 alLeuAlaLeuGIuGIu 521
::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5380 TCCTGACCTTAGACGAT 5364

```



```

51 GlyAlaThrGlyGlnArgGlyTyrProThrThrArgArgThrValAs 67
|||||
2310 GGTGCCACAGGGGACACACATATCCATGAGCTACCCGAGAACAGTGTGA 2359
67 PheGlyValGlyArgValThrHisSerPheLeuValIleProGluCysP 84
|||||
2360 CTTGGGAGTGGGACGGGTAAACCACTCGTTCTGTGTCATACCTGAGTGC 2409
84 roValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
|||||
2410 CAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGAGCAACAATT 2459
101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh 117
|||||
2460 TCTTTGACAAAGGGAACACAGAGTGTCTGCAATAACAAACCTATCAC 2509
117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnI 134
|||||
2510 TGTGTTGACCTCCCAATTAGATGACCAATATCGACTATATCTCTCCCTAG 2559
134 alLysProAspGlnAspIleGlnSerTyrPheGluGlnPheProGlnAla 150
|||||
2560 TAAAGCTGATCAAAATATCAATTCTGTGTAACAGATTCCCAAGCC 2609
151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVal 167
|||||
2610 TGGGCAGAAACCCGAGGATGGGTTGGCAAGCAAGTTCGCCCAACAGT 2659
167 ILeuGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrProL 184
|||||
2660 TATTCAACTAGAACGCCAGTCCACACACATGTCAGTCAAGACAGTACCCT 2709
184 euSerArgGluAlaArgGluGlyIleTyrProHisValGlnArgLeuIle 200
|||||
2710 TGAGTAAAGAGCTCAAGAGGAAATCGGCCGATGTCCAAAGATTAACT 2759
201 GlnGlnGlyIleLeuValProValGlnSerProTyrPAsnThrProLeu 217
|||||
2760 CAACAGGGCAATCCTAGTCTCTGTCATCTCCCTCGAATACCTCCCTGCT 2809
217 uProValArgLysBProGlyThrAsnAspTyrArgProValGlnAspLeu 234
|||||
2810 ACCGGTAAAGAACCTGGAGCTATGACTATCGACCAACAGACAGACTTGA 2859
234 rGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
|||||
2860 GAGAGGTCATTAAGGGGTGAGATATACACCCAAACAGTCCGGAACCT 2909
251 TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrVal 267
|||||
2910 TATAACCTCTGTGTGCTCTCCACCCCAAGGAGGTGATATACAGTATT 2959
267 uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
|||||
2960 GGACTTAAGAGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCAAAC 3009
284 roLeuPheAlaPheGluTyrPArgAspProGlyThrGlyArgThrGlyGln 300
|||||
3010 CACTTTTGGCTTCGAATGGAGATCCAGGTACGGAGAACCGGAGCAG 3059
301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
|||||
3060 CTCACCTGGAGCCGACCTCCCAAGGGTTCAAGAACTCCCGGACCACTTT 3109
317 eaSpGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
|||||
3110 TCACAGAGCCCTTACACAGAGACTGGCCCACTTAGGATCCACACACCTC 3159
334 IValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThr 350
|||||
3160 AGGTAGACCTCTCCAGTACGTGATGACCTGCTTGGCGGAGCACC 3209
351 LysGlnAspCysLeuGluGlyThrLysAlaLeuLeuGluLeuSerAs 367
|||||
3210 AAACGAGACTCTTAGAAGGACACAGACACTGCTGGAATGTCTGA 3259
367 PheGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGlu 384
|||||
3260 CCTAGGCTACAGAGCCTCTGCTAAGAGGCCCAATTTGGACAGAGAGG 3309
384 aThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrPheThrGlu 400
|||||
3310 TAACTACTGGGGTACAGTTTGGCGGAGCGGACGCATGGCTGACGGAG 3359
401 AlaArgLysLysThrValAlaGlnIleProAlaProThrThrAlaLysG 417
|||||
3360 GCACGGAAGAAACTGTAGTCAGATACCGGCCCAACACAGCCAAACA 3409
417 nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTyrIleProG 434
|||||
3410 AATGAGAGATTGTTGGGACAGCTGGATTTTGCAGACTGTGGATCCCGG 3459
434 IYPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGly 450
|||||
3460 GGTTCGACACTTAGACGCCCACTCTACCCGCTAACCAAGAAAGAGG 3509
451 GlyPheSerTyrAlaProGluHisGlnLysAlaPheAspAlaIleLysLy 467
|||||
3510 GAATTCCTCTGGGCTCTGACGACCAAGAGCATTGATGCTATCAAAA 3559
467 sAlaIleLeuSerAlaProAlaLeuAlaLeuProAspValThrLysPro 484
|||||
3560 GGCCCTGCTGAGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3609
484 heThrLeuTyrValAspGluArgGlyGlyValAlaArgLysValLeuThr 500
|||||
3610 TTACCTTTATGTGATGAGCGCTAAGGAGTGGCCGGGAGATTTAACC 3659
501 GlnThrLeuGlyProTyrPArgProValAlaTyrLysSerLysLysLe 517
|||||
3660 CAACCTTAGACCATGAGAGAACCTGTGCGCTACCTGTCAAGAACCT 3709
517 uAspProValAlaSerGlyTyrProValCysLeuLysAlaIleAlaIle 534
|||||
3710 CGATCTGTAGCCAGTGTGGCCCATATGCTGAGAGGCTATCGACGCTG 3759
534 aAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyLysnIle 550
|||||
3760 TGGCCATACGTGTCAAGAGCTGACAAATTGACTTTGGGACAGAAATATA 3809
551 ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAs 567
|||||
3810 ACTGTATAGCCCCCATGCTGATGGAGAACATGCTTGGCAGGCCCAAGA 3859
567 PArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuTyr 584
|||||
3860 CCGATGATGACCAACGCCCGCATGACCCACTATCAAAAGCTCTCTCTCA 3909
584 hrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
|||||
3910 CAGAGAGGTACTCTGCTGCCACACAGCCGCTCAACCTGTCCACTCTT 3959
601 LeuProGluGluThrAspGluProValThrHisAspCysHisIleLeuLe 617
|||||
3960 CTGCTGAAAGACTGATGAACCAAGTACATCAGATTTGCCATCAACTATT 4009
617 uIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThr 634
|||||
4010 GATTGAGGAGACTGGGGTCCGCAAGAGACTTACAGACATACCGCTAGCT 4059
634 LysGluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650
|||||
4060 GAGAGTGTCAACCTGCTTCACTGACGAGAACACTATGTGGTGAAGGT 4109
651 LysArgMetAlaGlyAlaIleValValAspGlyThrArgThrIleTrpAl 667
|||||

```

```

4110 AAGAGATGCTGGGCGCGGTGTGTGACGGAGCCGACGATCTGGGC 4159
667 aSerSerLeuProGluGlyThrSerAlaGlnValAlaGluLeuMetAla 684
4160 CAGACGCTCGCGGAGAGAACTTCAGCACAAGAGCTGAGCTCATGCGCC 4209
684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
4210 TCACGCAAGCTTTCGGCTGCGCCAGAGGAATCCATTAACATTATTATAC 4259
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisLeuAlaIleTyr 717
4260 GACACGAGGTATGCCCTTTCGACATGCACACGTACATGGGCGCATCTATA 4309
717 sGluArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 734
4310 ACAAAGGGGTTGCTTACCTCAGCAGGAGGGAATTAAGAAACAAGAGG 4359
734 IuileuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAlaIle 750
4360 AAATTCTAGGCTATTAGAAAGCCGTACATTACCAAAAAGGCTAGCTATT 4409
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgLys 767
4410 ATACACTGCTCTGCACATCAGAAAGCTAAAGATCTCATATCCAGAGAAA 4459
767 nGluMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 784
4460 CCAGATGCTCAGCCGGGTGCCAAGCAGGAGCCAGGGTGTTAACCTTC 4509
784 euProIleIleGluThrProLysAlaProGluProArgArgGlnTyrThr 800
4510 TGCCATATATAGAAATGCCCAAGCCCAAGAACCCAGACGACATACACC 4559
801 LeuGluAspTrpGlnGluIleLysLysIleAspIlePheSerGluThr 817
4560 CTAGAGAGCTGCGAAGATAAAAAGATAGACCAAGTCTCTGTGAGCTCC 4609
817 oGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLys 834
4610 GGAGGGAGCTGTATACCTCAGATGGGAGGAATCTGCCCCACAAG 4659
834 IuGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
4660 AAGGTTAGAAATATGTCACAGATACATGCTTAACCCACCTAGAACT 4709
851 LysHisLeuGlnGlnLeuValArgThrSerProGlyHisValLeuArgLe 867
4710 AAACACCTGCAGAGTTGTGTCAGAAACATCCCTTATCATGTTCTGAGGCT 4759
867 uProGlyValAlaAspSerValLysHisCysValProCysGlnLeu 884
4760 ACCAGAGTGGCTGACTGCTGCTCAAAACATTGTGTGCGCTGCCAGCTGG 4809
884 aLAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLys 900
4810 TTAATGCTAATCTCCAGAAATGCTCCAGAGGAGAGACTAAGGGGAAAG 4859
901 HisProGlyAlaHisThrGluValAspPheThrGluValLysProAla 917
4860 CACCAGGCGCTCAGCTGGAGAGTGCATCTCACTAGGTAAGGCGGCTAA 4909
917 sTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 934
4910 ATACGGAACAATATCATTTGTTTGTGACACCTTTCCAGAGATGGG 4959
934 aGluAlaTyrProThrLysLysGluThrSerThrValAlaLysLys 950
4960 TAGAGGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGAATAAAAA 5009
951 IleLeuGluGluIlePheProArgPheGlyIleProLysValIleGly 967
5010 ATACTGGAAGAAATTTTCCAGATTTCGAATACCTTAAGTAAATAGGCTC 5059

```

```

967 rAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysI 984
5060 AGACATGTCGCGAGCTTTGTGCCCCAGGTAGTGACGAGCTGGCCAAAG 5109
984 IeLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 1000
5110 TATTGGGAGTATGATGGAAATGTCATTTGTCATACAGACCCCAAGCTCA 5159
1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLys 1017
5160 GCACAGGTAGAGAGATGAATAGAACCATTAAGAACCCCTTACTTAATT 5209
1017 uThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1034
5210 GACCGGAGAGACTGGCGTTAATGATGTGATGATGTCCTGCGCTTGTGCG 5259
1034 euPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProGly 1050
5260 TTTTATGAGGTAGGAACACCCCTGCAGAGTTTGGCTGACCCCTATGAA 5309
1051 LeuLeuTyrGlyLysProProProLeuAlaGluIleAlaPheAlaHis 1067
5310 TTACTCTACGGGGAGACCCCCCATTTGTAAGAAATGCTTCTGTACATAG 5359
1067 rAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1084
5360 TGCTGACGTGCTGCTTCCAGGCTTGTGTTCTCTAGGTCAGAGGCACTTG 5409
1084 IuThrValArgGlnArgAlaTyrLysGlnLeuArgLysAlaTyrSer 1099
5410 AGTGGGTGAGACAAACGAGGTGAGCAACTCCGGGAGGCGCTACTAGGA 5459
1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerVal 1116
5460 GGAGGAGACTGTCAGATCCCATGCTGTTTCCAGTGGAGATTCACATCTA 5509
1116 rValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyPro 1133
5510 CGTTAGACGCGCACCGTGCAGAAACCTCGAGATCGGTGGAAGGCGCTT 5559
1133 yLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIlePro 1149
5560 ATCTGCTACTTTTGACCAACCAACGCGCTGTGAAGTGCAGAAAGAACTCC 5609
1150 ThrThrIleHisAlaSerHisValLysProAlaProProAspSerG 1166
5610 ACCTGATTCATCATCCACAGTTAAACCGGCGCCACCTCCGATTCGGG 5659
1166 yTrpArgAlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArg 1183
5660 GTGGAAGAGCGGAAAGAGCTGAATAATCCCTTAAGCTTGGCTCCAGTGG 5709
1183 euValProTyrSerAsnAsnAsn 1190
5710 TGTGCTCTTACTCTGTCAATAATAC 5732

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

seq_documentation_block:

; Sequence 1, Application US/08766528

; Patent No. 6190861

; GENERAL INFORMATION:

; APPLICANT: Jay A. Fishman

; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESS: LAHIVE & COCKFIELD, LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

```

ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

```

```

alignment_scores:
Quality: 5637.50      Length: 1127
Ratio: 5.061          Gaps: 3
Percent Similarity: 98.846      Percent Identity: 96.628

```

alignment block:
US-09-171-553B-5 x US-08-766-528-1 ...

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1  GLYARGATGGLYSERAPPROLEUPROGLUPROARVALTHIRLEULYSVA 17
   |||||||
4687 GGGAGACGGGGTTCGGACCCCTCCCGAGCCCGGGGTAACTTTGAAAGGT 4736
17  |GLUGLYG|NPROVALGIUPHEULVALSPRTHGLY|AGL|UHSERY 34
   |||||||
4737 GGAAGGGCAACCAAGTGTCTGCTGATTGATACCGGAGCGAAACATTCAG 4786
34  |ALLEULEG|NPROLEUG|LYSLEULYSGL|ULYS|SER|TR|VAL|MET 50
   |||||||
4787 TGTCTACTACAGCCATTTAGGAATAAAGATAAATAAATCCTGGGTGATG 4836
51  |GLY|ATPR|G|LY|N|ARG|GL|N|TYR|P|RO|TR|P|TH|RA|GT|A|G|TH|VAL|AS 67
   |||||||
4837 GGTGC.ACAGGGCAACAACAGTATCCATGACTACCCGAAG.ACAGTTGA 4884
67  |PLEUG|LY|VAL|GLY|AR|VAL|TH|RH|SER|PHE|ULVAL|LE|PRO|GL|UC|YSP 84
   |||||||
4885 CTGCGAGTGGGACGGGTAACCCACCTGCTTCTGTCATACCTGAGAGGCC 4934
84  |ROVAL|PROLEULEG|LY|ARG|ASPLEULE|U|H|TH|LYS|MET|GLY|AL|GL|ILE 100
   |||||||
4935 CAGACCCCTCTTAVGGTAGAGACTATTATTGACCAAGATGGGAGCACAATTT 4984
101 |SER|PHE|GL|UG|N|GLY|AR|PRO|GL|U|VAL|SER|VAL|ASN|N|LYS|PRO|ILE|TH 117
   |||||||
4985 TCTTTTAAACAAGGAAACCAAGAGTGTGCAATATACAAACCATTCATCAC 5034
117 |VAL|LEU|TH|LEU|GL|N|LEU|ASN|ASP|GL|U|TYR|ARG|LEU|TYR|SER|PRO|GL|N 134
   |||||||
5035 TGTGTGACCCCTCCATTTAGATGAGATATACGACTATCTACCTCCCTAG 5084
134 |ALY|SP|RO|ASP|GL|N|ASP|ILE|GL|N|SER|TR|LEU|GL|N|PHE|PRO|GL|N|ALA 150
   |||||||

```

```

5085 TAAAGCCTGATCAAAATATACAAATCTGTGGACAGGTTTCCCAAGCC 5134
151 |TRP|AL|GL|U|TH|RH|AL|GLY|MET|GLY|LEU|AL|ALY|GL|N|VAL|PRO|PRO|GL|N|VA 167
   |||||||
5135 TGGGCGAAGAAACCCGAGGATGGGTTGGCAACCAAGTTCCCAACAGT 5184
167 |ILE|GL|N|LEU|LYS|AL|SER|AL|ATH|RH|PRO|VAL|SER|VAL|ARG|GL|N|TYR|PRO|L 184
   |||||||
5185 TATTCACCTGAGGCCAGTCGCACACCAAGTGTGATGACAGACTACCCCT 5234
184 |EUS|ER|ATG|GL|U|AL|AR|GL|UG|LY|ILE|TR|PRO|HIS|VAL|GL|N|ARG|LEU|ILE 200
   |||||||
5235 TGAATTAAGAAAGCTCAAGAAAGAAATTCGGCGGATGTCCAAATTAATTC 5284
201 |GL|N|GL|N|LY|LEU|VAL|PRO|VAL|GL|N|SER|PRO|TR|P|AS|N|TH|P|RO|LEU|E 217
   |||||||
5285 CAACAGGGCATCTCTAGTCTGTCCATCTCCCTGGAAATATCTCCCTGCT 5334
217 |UPR|OVAL|ARG|LYS|PRO|GL|Y|TH|RH|ASN|ASP|TYR|ARG|PRO|VAL|GL|N|ASP|LEU|A 234
   |||||||
5335 ACCGGTTAGAAAGCCTGGGACTAATGACTATGACACAGTACAGGACTTGA 5384
234 |RGL|U|VAL|ASN|LYS|ARG|VAL|GL|N|ASP|ILE|HIS|PRO|TH|VAL|PRO|ASN|PRO 250
   |||||||
5385 GAGAGGTCAATTAACGGGTGCAGAGATATACACCAACAGTCCCAACCTCT 5434
251 |TYR|ASN|LEU|LEU|SER|AL|ALEU|PRO|PRO|GL|U|ARG|ASN|TR|P|TH|VAL|LE 267
   |||||||
5435 TATTAACCTCTTGTGTCTCTCCACCCCAACGAGGCTGTATATACATATT 5484
267 |UASP|LEU|LYS|ASP|ALA|PHE|PHE|CYS|LEU|AR|GLY|N|HIS|PRO|TH|SER|GL|N|P 284
   |||||||
5485 GGACTTAAGAGATGCTTCTTCTGCTGAGATATACACCCACAGCCAGCCAAC 5534
284 |ROLEU|PHE|AL|PHE|GL|U|TR|P|ARG|ASP|PRO|GL|Y|TH|GLY|ARG|TH|GLY|GL|N 300
   |||||||
5535 CACTTTTTCCTTCGAATGAGATCCAGATCCAGTACGGGAAACCGGGCAG 5584
301 |LEU|TH|TR|P|TH|ARG|LEU|PRO|GL|N|GLY|PHE|LYS|ASN|SER|PRO|TH|ILE|PH 317
   |||||||
5585 CTACACCTGCACCCGACTGCTCCCAAGGTTTCAAGACTCCCAACCATCTT 5634
317 |EASP|GL|U|AL|ALEU|HIS|ARG|ASP|LEU|AL|ASN|PHE|ARG|ILE|GL|N|HIS|PRO|G 334
   |||||||
5635 TGACGAAGCCCTACACAGAGCCTGGCCAACTTCAGAGTCCAAACACCCTC 5684
334 |IN|VAL|TH|LEU|LEU|GL|N|TYR|VAL|ASP|ASP|LEU|LEU|AL|GLY|AL|ATH|RH 350
   |||||||
5685 AGGTGACCTCTCCCTCCAGTACGTGGATGACCTCTTGTGCGGAGCCACCC 5734
351 |LYS|GL|ASP|CYS|LEU|GL|UG|LY|TH|RH|LYS|AL|ALEU|LEU|GL|N|LEU|SER|AS 367
   |||||||
5735 AATCAGAGACTGCTTTAGAGGCAACAGAGGCACTACTGCTGGAATTTGCTGA 5784
367 |PLEU|GLY|TYR|ARG|AL|SER|AL|LYS|LYS|AL|GL|N|ILE|CYS|ARG|ARG|GLY|U 384
   |||||||
5785 CTAAGGCTACAGAGCCTCTGCTAAGAGGCCCAAGATTTCAGAGAGAGAGG 5834
384 |AL|TH|TYR|LEU|GLY|TYR|SER|LEU|AR|GLY|GLY|GL|N|ARG|TR|LEU|TH|GLU 400
   |||||||
5835 TAACATACCTTGGGTACAGATTACGGGAGCGGACGAGTGCCTGACGAG 5884
401 |AL|AR|GLY|LYS|TR|VAL|VAL|GL|N|LEU|PRO|AL|PRO|TH|TH|AL|ALY|GL| 417
   |||||||
5885 GCACGGAAGAAACCTGTAGTCCAGATACCGGCGCCCAACACGCAACCA 5934
417 |NVAL|ARG|U|PHE|LEU|GLY|TH|RAL|GLY|PHE|CYS|ARG|LEU|TR|P|ILE|PRO|G 434
   |||||||
5935 AATGAGAGAGTTTGGGACAGCTGAGANTTTGACAGACTGTGATCCCGG 5984
434 |LYPHE|AL|ATH|RH|LEU|AL|AL|PRO|LEU|TYR|PRO|LEU|TH|LYS|GLY|ULYS|GLY 450
   |||||||
5985 GGTTCGACCTTAGACGCCACCTACCCGCTAACCAAGAAAGAGG 6034

```

451 GlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaIleuLys 467
|||||
6035 GAATCTCTGGGCTCTGAGCACCAGAGCATTTGATGCTATCCAAAA 6084
467 sAlaIleuSerAlaProAlaIleuAlaIleuProAspValThrLysPro 484
|||||
6085 GGGCTGCTGAGCGCAGCTGCTGGCCCTCCCTGACCTAGCTAAACCT 6134
484 heThrLeuTrpValAspGluArgLysGlyValAlaArgGlyValLeuThr 500
|||||
6135 TTACCTTTATGTGATGAGCGTAAGSAGTAGGCCGGGAGTTTAAAC 6184
501 GlnThrLeuGlyProTrpArgArgProValAlaIleuSerLysLys 517
|||||
6185 CAACCCCTAGGAGCCATGGAGAAAGCTGTCCCTACCTGTCATAAGAACT 6234
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaVal 534
|||||
6235 CGATCTGTAGCCAGTGGTGGCCCATATGCTGAGAGCTATCCGAGCTG 6284
534 aAlaIleuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
|||||
6285 TGGCCATCTGCTCAAGACGCTGACAAATTGACTTTGGACAAAGATAT 6334
550 eThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProPro 567
|||||
6335 AACTGTATAGCCCCCATGCAATTGGAGAACATCTTCGCGACCCCGAC 6384
567 sPArgTrpMetThrAsnAlaArgMetThrHisTrpGlnSerLeuLeu 583
|||||
6385 ACCGATGATGACCAACGCCCATGACCCATCAAGCTGCTTCTC 6434
584 ThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLe 600
|||||
6435 ACAGAGAGGGTCACTGCTGCCCAACCAACCGCTCAACCGTGCACACT 6484
600 uLeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeu 617
|||||
6485 TCTGCTGAGAGACTGATGAAACAGTACTGATGATGCTCAACTCAACT 6534
617 euIleGluGluThrGlyValAlaArgLysAspLeuThrAspIleProLeuThr 633
|||||
6535 TGATTGAGAGAGCTGGGGTCCGCAAGACCTTACAGACATACCGCTGACT 6584
634 GlyGluValLeuThrTrpPheThrAspGlySerSerTrpValValGluG 650
|||||
6585 GGAGAGTGTCTAACTGTCTACTGACGAGAACACTATGTGTGGAAGG 6634
650 yLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIleTrpA 667
|||||
6635 TAAAGAGATGGCTGGGGCGGGCTGTGAGCGGAGCCCGCAGATCTGGG 6684
667 laSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAla 683
|||||
6685 CCAGCAGCTGCCGCGGAGAACTTACGACAAAGAGCTGAGCTCATGGCC 6734
684 LeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleuThr 700
|||||
6735 CTCACCAAGCTTGGCGCTGGCGGAGGAAATCCATTAACATTTATAC 6784
700 rAspSerArgTrpAlaPheAlaThrAlaHisValHisGlyAlaIleTrpL 717
|||||
6785 GGCACAGAGGTATGCTTTGGCAGCTCCACAGCATGAGGGGCATCTATA 6834
717 ySGLnArgGlyLeuLeuThrSerAlaGlyArgGluLysLysGlu 733
|||||
6835 AACAAAGGGGGTGTCTTACTCTCAGCGAGGGGAAATTAAGAAACAAAGAG 6884
734 GluIleuSerLeuGluAlaLeuHisIleuProLysArgLeuAlaIle 750
|||||
6885 GAAATTTCTAAGCTATTAGAACCGGTACATTTCACAAAAGGCTACTAT 6934

750 eIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyA 767
|||||
6935 TATACACTGCTCTGGACATCAGAAAGCTTAAAGTTCATATCCAGAGGAA 6984
767 sngIleuAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 783
|||||
6985 ACCAGATGGCTGACCGGTTGGCCACAGCAGCAGCCAGGGTGTAACTT 7034
784 LeuProIleIleGluThrProLysAlaProGluProArgArgIleuThr 800
|||||
7035 CTCCTATATATAGAAATGCCCCAAAGCCCGAAGCCCAAGACGACAGTACAC 7084
800 rLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrP 817
|||||
7085 CTTAGAAAGACTGGCAGAGATGATAAAAAGATAGACCA.TTCTGTGAGACTC 7133
817 roGluGlyThrCysTrpThrSerTrpGlyLysGluIleuProHisLys 833
|||||
7134 CGGAGGAGACTGCTATACCTCAGATGGAGAAATCCGCCCAACAA 7183
834 GluGlyLeuGluTrpValGlnGlnIleHisArgLeuThrHisLeuGlyT 850
|||||
7184 GAAGGTTAGATATGTCTCAACAAAGATACATCGCTTACCCACTAGGAA 7233
850 hTrpHisLeuGlnGlnLeuValArgThrSerProTrpHisValLeuArg 866
|||||
7234 CTAACACCTGACAGCACTTGTGAGAACATCCCTTATCATGTCTGAGG 7283
867 LeuProGlyValAlaAspSerValValLysHisCysValProCysGlnLe 883
|||||
7284 CTACACAGAGTGGCTGACTCGTGTCAAAACATTTGTGCTGCCACGCT 7333
883 uValAsnAlaAspProSerArgIleProProGlyLysArgLeuArgLys 900
|||||
7334 GGTAAATGCTATCTCTCCAGAAATGCTCCAGGAAAGACTAAGSGGAA 7383
900 eHisProGlyAlaHisIleTrpGluValAspPheThrGluValLysProAla 916
|||||
7384 GCCACCGAGCGCTCACTGGGAAAGTGAAGTCACTGATGAGTAAAGCCGCT 7433
917 LysTrpGlyAsnLysTrpLeuLeuValPheValAspThrPheSerGlyTr 933
|||||
7434 AAATATGGAACAATATCTATGTTGTTTGTAGACACTTTTCAGATG 7483
933 pValGluAlaTrpProThrLysLysGluThrSerThrValValAlaLysL 950
|||||
7484 GGTAGAGCTTATCTACTAGAAAGAACTTCAACCGTGTAGCTTAAA 7533
950 ySILeGluGluGluIlePheProArgPheGlyIleProLysValIleGly 966
|||||
7534 AAATACTGGAAGAAATTTTCCAGATTTGGAATACCTTAAGTAAATAGG 7583
967 SerAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
|||||
7584 TCAGCAATGTGTCAGCTTTTGTTCGCCAGGTATGTCAGGACGGCCAA 7633
983 sILeGluGlyIleAspTrpLysLeuHisCysAlaTrpArgProGlnSer 1000
|||||
7634 GATATTGGGATGTTGATTGAAACTGCAATGTGCATACAGACCCCAAGCT 7683
1000 eArgGlyValGluArgMetAsnArgThrIleLysGluThrLeuThrLys 1016
|||||
7684 CAGCAGAGTAGAGAGATGATGAAACCAATTAAGAGACCTTACTATAA 7733
7734 TTGACCGGGAGAGACTGGCGTTATGATGATGATGATGATGATGATGAT 7783
1017 LeuThrTrpGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
|||||
1033 lLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTrpG 1050
|||||
7784 GCTTTTATAGGTTAGGAACACCCCTGACAGCTTTGGGTGACCCCTATG 7833
1050 lLeuLeuTrpGlyLysProProLeuAlaGluIleAlaPheAlaHis 1066

```

7834  AATTACTCTACGGGGGAGCCCCCCCCCATGTGAGAAATGCTTCGTGATCAT 7883
1067  Ser11aspVal11leuSerGlnProleuPheSerArg1eulysAla1e 1083
7884  AGTCTGATGTGCTTGCTTCCAGCCCTTGTCTCTAGGCTCAAGGCACT 7933
1083  uGU1rPvalArgGlnArgAla1rPlysGlnLeuArgGlnAla1rYser 1099
7934  TGAATGGGTGAGACACAGAGCGTGGAGGCACTCCGGAGGCTTACTACAG 7983
1100  ..GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerVal 1115
7984  GAGGAGGAGACTTCGACATGCCACATGTTTCCAAAGTGGGAGATTCAATC 8033
1116  TyrValArgArgHisArgAlaGln1Asn 1124
8034  TACCTTAGACGCCACCGTGCAGGAAC 8060

seq_name: /cgn2_6/prodata/2/1na/6B.COMB.seq:US-08-766-528-2

seq_documentation_block:
: Sequence 2, Application US/08766528
: Patent No. 6190861
: GENERAL INFORMATION:
: APPLICANT: Jay A. Fishman
: TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 74
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 60 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/766,528
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/572,645
: FILING DATE: 14-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Louis Myers
: REGISTRATION NUMBER: 35,965
: REFERENCE/DOCKET NUMBER: MCP-038CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 733 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-08-766-528-2

alignment_scores:
Quality: 4489.50 Length: 1191
Ratio: 5.033 Gaps: 2
Percent Similarity: 74.895 Percent Identity: 74.643

alignment_block:
US-09-171-553B-5 x US-08-766-528-2 ..

Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333

```

1	GLYAPRGRGGLSERAPSPROLEUPROGLUPROARGVALTHHLEUYSVA	17
2173	GGGAGAGGGGGTTGGAGACCCCTCCCGAGCCAGGGTAACTTGAAGCT	2222
17	IGLUGLGNPROVALGIDUPHELEUVALASPTHRLGLYALGIDHISERV	34
2223	GGAGGGGCAACCGATGGAGTTCCTGGTTGATACCGGAGCGGAGCAATTCAG	2272
34	ALLEULEUGINPROLEUGILYLSLEULEYSGULYLSYSSECTRPALMET	50
2273	TGCTGCTACACCAATTBAGGAAAACTBAAAGAAAAAAATTCGGGGTGTWG	2322
51	GLYALATHRLGLSINARGINTYRPROTPHTRHTRARGATGYTHVALS	67
2323	GGTCCACAGGGGCAAGCGGCGATTCATCGATGACTACCCGAAAGACCTTGA	2372
67	PLEUGLYVALGLYARVALTHHISERPHLEUVALIIEPROGLUCYSP	84
2373	CTTGGGGGTGGGAGGGGTAAACCACACGCTTCTGGTCAATCCCTGAGTGCC	2422
84	ROYALPROLEUGILYARGASPLEULEUTHLYSMEGLYALGIDHILE	100
2423	CAGTACCCCTCTTAGGTAGAGACTTACGTACCAAGATGGGAGCTCAATT	2472
101	SERPHLEUGINGLYARPROGLUVALSERVALASNAULYSPROIETRH	117
2473	TCATTTGGAAACAGGAGACCAAGACAGTGTGTAATACAAACCCATCAC	2522
117	VALLEUTHRLLEUGINLEUASPSGLTYTHRGLYEURYSERPROGLIN	134
2523	TGTGTTGACCCCTCCATTATGATGATGAATATCGACTTATATCTCCCAAG	2572
134	ALLYSPROASPSGLNASPILEGINSERTTRPLEUGINLPHENPROGLINAL	150
2573	TAAAGCGCTATCAAGTATATACAGCTCCGTGGTGGAGCAGATTCCCAAGCC	2622
151	TRPALGIDHTRALGIDYMETGLYLEUALALYSGINVALPROPROGLINVA	167
2623	TGGGCGAAACCGCAGGGGTGGGTTGGCAAAAGCAAGTCCCCCAAGGT	2672
167	LIIEGLINLEULYSALASERIALATHPROVALSERVALARGINLYRPROL	184
2673	TATTCACATGGAAGGCCAGTGCATACACAGTATCAGTACAGACAGTACCCT	2722
2723	TGAGTAGAAGAGGCTCGAGAAAGAAATTTGGCCGATGTGTTCAAAAGATTATAC	2772
201	GLINGLIGLYILELEUVALPROVALGIDSEXPROTFRPSANTHRPROLEUE	217
2773	CAACAGGCGATCTTAGTTCCTGTCCCAATCCCTTGGANATCTCCCTGCT	2822
217	UPROVALARGLYSPROGLYTHHRASNASEPYRARGPROVALGIDNASPLEUA	234
2823	ACCGGTTAGGAAGCCTCGGAGCAAAATGATTATTCGACCAAGTACAGACTTGA	2872
234	IRGLIIVALASNLYSARGVALGIDNASPILEHISPROTHRVALPROASNPRO	250
2873	GGAGGGTCAATTAAGAGGGGTGGAGGACATACCCCAAGCGTCCGAAACCTC	2922
251	TYRASNLEULEUSERVALALEUPROGLIUAARGSNTRPYRTHRYALLE	267
2923	TATTAACGCTTGGAGGCCCTCCGCCCTGGAACGGAACTGGTACACAGATT	2972
267	UNSPLEULEYSPALAPHEPIECYSLLEUARGLEUHSISPROTHRSEGLNP	284
2973	GGACTTTAAAGATGGCTTTCTCTCCGAGATTTACACCCCACTAAGCCAAC	3022
284	ROLEUPHALAPHEGLIUTRPRARGASPROGLYTHRLGLYARGTHRLYGIN	300
3023	CACTTTAACTCTCGAATGGGAGAGATCCAGGTACGGGAAACACCGGGGCAAG	3072

301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePhe 317
|||||
3073 CTCACCTGGACCCGACTGCCCCAGAGGCTTCAAGACTCCCCGCCATCTT 3122
317 eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
|||||
3123 TGAAGAGCCCTACACAGGAGCCTGGGCCACTTCAGATCCACACCCTTC 3172
334 InValThrLeuLeuGlnThrValAspAspLeuLeuAlaGlyAlaThr 350
|||||
3173 AGGTGACCTCTCCACGTACGTGATGACCTCTCTGGCGGAGGACACC 3222
351 LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGlnLeuSerAs 367
|||||
3223 AACAGAGACTGCTTAGAAGGTACAGAGCCACTACTGCTGGAATTGCTGA 3272
367 PLeuGlyThrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluV 384
|||||
3273 CCTAGGCTACAGAGCCTCTGCTAGAGAGCCAGATTTCAGAGAGAGAG 3322
384 aLThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400
|||||
3323 TAACTACTTGGGGTACAGTTTGGGGGCGGCGACGATGCTGACGAG 3372
401 AlaArgLysLysThrValAlaGlnIleProAlaProThrThrAlaLysG 417
|||||
3373 GCACGGAGAAAGAACTGTAGTCCAGATACCGGCCCAACACAGCAACACA 3422
417 nValArgGluPheLeuGlnThrAlaGlyPheCysArgLeuTrpIleProG 434
|||||
3423 AGTGAGAGAGTTTGGGGACAGCTGGATTTCGACACTGTGATCCCGG 3472
434 LyrPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLysGly 450
|||||
3473 GGTTCGGACCTTAGCAGCCCTACCCGCTAACCCCAAGAAAGAGGG 3522
451 GlyPheSerTrpAlaProGlnHisGlnLysAlaPheAspAlaIleLysL 467
|||||
3523 GGTTCG..... 3527
467 sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysPro 484
3527 3527
484 nethrLeuTyrValAspGluArgLysGlyValAlaIaArgGlyValLeuThr 500
3527 3527
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLe 517
3527 3527
517 uAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaIaV 534
3527 3527
534 aAlaIaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle 550
3527 3527
551 ThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProProAs 567
3527 3527
567 PArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuT 584
3527 3527
584 hngIuArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600
3527 3527
601 LeuProGlnGluThrAspGluProValThrHisAspCysHisGlnLeuLe 617
3527 3527
617 uIleGlnGluThrArgValArgLysAspLeuThrAspIleProLeuThrG 634
3527 3527
634 LysIuValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 650
3527 3527
651 LysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIleTrpAl 667
3527 3527
667 aSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGluLeuMetAlaL 684
3527 3527
684 euThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
3527 3527
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrI 717
3527 3527
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluG 734
3528 3527
734 LuIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
3565 AAATTCTAAGCCATATAGAACCTTACATTGTGCCAAAAGGCTAGCTATT 3614
751 PLeuHisCysProGlnHisGlnLysAlaLysAspLeuIleSerArgGlyLys 767
|||||
3615 ATCACTGTCTGTCGACATCAGAAAGCCAAAGATCTCATATGTCAGAGGAA 3664
767 ngIleMetAlaAspArgValAlaLysGlnAlaIaGlnAlaValAsnLeuL 784
|||||
3665 CCAGATGGCTGACCGGGTGGCCAGCAGGACGCCCGGCTTTAACTTC 3714
784 euProIleIleGlnThrProLysAlaProGluProArgArgGlnTyrThr 800
|||||
3715 TGCCCTAATATAGAAACCCCAAGCCCAAGCCAGACAGACAGTACACC 3764
801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrPr 817
|||||
3765 CTGAGAGACTGGCCAGAGATAAAGAAAGATAGACAGACTCTCTGAGACTCC 3814
817 OGlnGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysG 834
|||||
3815 GGAAGGGACCTGCTATACCTCATATGGAAGAAATCTGCCCAAAAG 3864
834 LuGlyLeuGlnTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThr 850
|||||
3865 AAGGGTAGAATATGTCCAAACAGATACATGCTTAAACCCACCTAGGAAC 3914
851 LysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLe 867
|||||
3915 AACACCTGCAGCAGTGGTCAGAGACATCCCTTATCATGTTCTGAGGCT 3964
867 uProGlnValAlaAspSerValAlaLysHisCysValProCysGlnLeuV 884
|||||
3965 ACCAGAGTGGCTGACTGGTGGTCAAACTTTGTGGCCCTGCCACACTGG 4014
884 aLAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLysSer 900
|||||
4015 TTATGCTATCTCTCCAGATATCTCCAGAAAGAGACTAAGGGGAGAGC 4064
901 HisProGlyValaHisTrpGluValaAspPheThrGluValLysProAlaL 917
|||||

2491	GAAGAGTAAAGAACATGGTGGCCACACTCTTCTCGTGATACCTGAGTGC	2540
84	royalProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle	100
2541	CTGGCTCCCTCTGGAGCTGATCTGTAAACAAACTAAAGCGCTCAAGTT	2590
101	SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleTh	117
2591	CAATTTACTTTCAGAAAGTCCACCAATAGCTGGGGGAAAGCCCTTCG	2640
117	YalleuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGln	134
2641	TTGCCTTGTCTCAGCAGCGAAGAGACTACCGGTTGCATGAGAGCAAC	2690
134	alLysProAspGlnAspIleGlnSerTyrLeuGlnPheProGlnAla	150
2691	CCAAAGGTGCACGCCCTTTTA...GACTGGTAACGTGGTTCGCCAATGTC	2737
151	ThrAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGln	167
2738	TGGGCGGAACAAGCAGGATGGGGTTGGCTAAACAAGTGCCTCAGTGT	2787
167	LIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrPro	184
2788	GGTAGACTTTAAAGCTGATGCTACCCCATCTCAGTAAAGACAATACCCA	2837
184	euserArgGluAlaArgGlnGlyIleTyrProHisValGlnArgLeuIle	200
2838	TGAGCAGGAAGACTAAGAGGGGCATCCGACTCATATTTCGGAGGCTGTA	2887
201	GlnGlnGlyIleLeuValProValGlnSerProTyrAsnThrProLeu	217
2888	GACCAAGGAATTTTAGTGCCGTGCAGTCCGCCCTGGAACACTCCACTTT	2937
217	uProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeu	234
2938	GCCAGTAGCAAAAGCCAGGACCAATGATTATCGCCGGTACAAACCTCC	2987
234	rgsIuValAsnLysArgValGlnAspIleHisProThrValProAspPro	250
2988	GGGAAGTTAATAAAGGGCTCGGACATTTCACCCGTACAGTCCGAAACCG	3037
251	TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTyrThrValLe	267
3038	TACAACTTACTAAGCTCTCTCCACCCAGAGAACTGGTATACGGTCT	3087
267	uasgLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerInp	284
3088	GGATTTTAAAGATGCCCTCTTTTGGCTCGTGCACCCCAAGGATTAAC	3137
284	roLeuPheAlaPheGluTyrPargAspProGlyThrGlyArgThrGlyGln	300
3138	TTCTGTTTGGCTTTGAATGGAGAGACCCAGAGGGCGGACAGACTGGTCAA	3187
301	LeuThrTyrThrArgLeuProGlnGlyPheLysAsnSerProThrIlePhe	317
3188	TTAACTGGACTAGACTACACAGGGGTTCAAAAACTCCCTACCTGTT	3237
317	eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG	334
3238	TGCGAAGGCCCTCCATCGGATTTAGCACCTTTCGGCGCTCAAAACCCC	3287
334	lnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThr	350
3288	AGCTTACCTACTGCACTACAGTACATGATCTTTTAATCGCAGCGGCTCA	3337
351	LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGluLeuSerAs	367
3338	AAAGACTATGTCACAGAGGGGACTGAGAGACTCCCTACAGAACTGGGAA	3387
367	pleuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGlu	384
3388	TTTGGGGATTCAGGTTTCGGTTAAAGGCACAAATTTGTCAAACTTAGG	3437
384	alThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrPheLeuThrGlu	400
3438	TAATCTAATCTGGGGTACTTTCGAGGAGGTAAAGATGGCTCACTGA	3487
401	AlaArgLysThrValAlaGlnIleProAlaProThrThrAlaLysG	417
3488	GCCCGGAAAAAGCTGATGATGATTCACCGCCACACACACAGCGCA	3537
417	nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuThrIleProG	434
3538	GGTACGTAGTTTCTGGGAGCTGCTGGCTTTGTAGACTGTGATTCAG	3587
434	LyrPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLysG	450
3588	GCTTGGAGCCCTAGAGACACCCCTATATCCCTTACATGAGGGAAGATT	3637
451	GlyPheSerThrAlaProGluHisGlnLysAlaPheAspAlaIleLys	467
3638	CCCTTGAATGGAAGAAGACACCAAGAGCTTTTGAGGCTATCAATC	3687
467	sAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrLysProp	484
3688	GTCTCATATGACGTCCCGCTGGCTAGCATTTACAGACTTGAATGCT	3737
484	hethLeuTyrValAspGluArgGlyValAlaAlaArgGlyAlaLeuThr	500
3738	TGCTCTATATGTGGACGAGAGCGGGCATTCGACAGGGGGTGTGCA	3787
501	GloThrLeuGlyProThrPargProValAlaTyrLeuSerLysLysLe	517
3788	CAAGCACTAGACCCCTGGAAAGACCGGTAGGCTATCTGTCAAAAAACT	3837
517	uasPProValAlaSerGlyTyrProValLysLeuLysAlaIleAlaVal	534
3838	GGATCCCGCTTGCTAGTGGATGGGCCACATCTGAAAGCTATTTGACGAG	3887
534	alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIle	550
3888	TAGCCCTGCTGATCAAAAGATGCGATGAATGACATGGGAGACGAGCTG	3937
551	ThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProPropAs	567
3938	ACTGCTGTGGCCCTATGCGCTGGAGGATTTGGCGGACCCACCTGA	3987
567	PargTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeu	584
3988	TAGATGATACTAATTCGCCGATGACACACTATCAGACTCTGCTGCTAA	4037
584	hrgIuArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu	600
4038	ATGACCGGTAAACCTTTGCCCCCTCGCATTTCTCAACCCAGCCACCTC	4087
601	LeuProGluGluThrAspGluPargValThrHisAspCysHisGlnLeuLe	617
4088	CTCCCTTAAAGATATTCGCTCCACCAATCATGATGACAGACTCT	4137
617	uIleGlnGluThrGlyValArgLysAspLeuThrAspIleProLeuThrG	634
4138	GGCTGAAGAAATTGGGACAGAAAGACTGTACTACCAACCTGGCTG	4187
634	lyGluValLeuThrTyrPheThrAspGlySerSerTyrValValGluGly	650
4188	GA...GCTCTACCTGGTACAGAGATGGACAGAGTTTCTGTGATGAGGGA	4234
651	LysArgMetAlaGlyAlaValAlaValAspGlyThrArgThrIleThrAl	667
4235	AAAGGAAGGGCTGAGCTGGGGTGGATGGAAAAAGGTAATTTGGG	4284
667	aserSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAla	684
4285	AAGTGCCTTCTCGAAGGAATTCGGCACAGAGGCTGAACCTATAGGC	4334


```

684 eUThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700
|||||
4335 TTACACAGGCGCTCCGACGAGGCGGAGATGATCATATTAATTAACACT 4384
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrThr 717
|||||
4385 GACAGCGCGCTATGCTTTCTTCTACTGACATATCATGCGGCGCATCTACAG 4434
717 sGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 734
|||||
4435 GCAGCGAGGGTTATTGACTTTCAGCCGCGTAAGACATCAAAACAAGAAG 4484
734 IuileuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIle 750
|||||
4485 AAATTTCTGCGCTGTAGAGCCATACAGCCCAAGAGGTACCATC 4534
751 IleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAs 767
|||||
4535 ATCCACTGTCGCGTCACCAAGAGAGAGACTTAGTGGCCAGAGGCGAA 4584
767 nGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeu 784
|||||
4585 CCGATGCGAGACTCAGTACAAACAGGTTCTCAGGGGCGCATGATCT 4634
784 euProIleIleGluThrPro...LysAlaProGluProArgArgGlnTyr 799
|||||
4635 TAACTGAAAGAAATCCGTCACAAAGCCCTAG... 4669
800 ThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluTh 816
|||||
4670 .....GATGAAACATGATATTAAGAATATTTGGAC 4704
816 rProGluGlyThrCysTyrThrSerTyrGlyLysGluIleuProHisL 833
|||||
4705 TAGTATCCCTCCCATCTTTTCGAAGAAATAAGACTTGACACCCG 4754
833 ySGlnGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849
|||||
4755 AGGAAGAAATTAATTTGGAAGAGATACACCAAGTTTACCCACCTGGGA 4804
850 ThrHisIleuGlnGlnIleuValArgThrSerProTyrHisValLeuAr 866
|||||
4805 GTTGAAGAAATGATGAGACTGATTAAGAAAGTCCCGTATCAAGTCCCTAA 4854
866 gluProGlyValAlaAspSerValValLysHisCysValProCysGln 883
|||||
4855 CTGGAAGTCACTGCTCAAAAGATTAATCACTCTCAAGCGTGTGAT 4904
883 euValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGly 899
|||||
4905 TCACTAATGCAACTAAACCTTCAAGAAAGACCTGGAAAGAGACAAAGCGGA 4954
900 SerHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAl 916
|||||
4955 GACCGTCTCGAGGTATTTGGAGGTGATTTTACTGAAGTTAAACCCGG 5004
916 aLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 933
|||||
5005 AATGATGATGATATAGTATCTGTTAGTATTTGTAGACACCTTTACAGAT 5054
933 rPValGlnAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949
|||||
5055 GGGTTGAAGCGCTCCCACTAAACCTGAGACTGCCAGATGGCGCAAG 5104
950 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleG 966
|||||
5105 AAGATTTTGAAGAAATCTCCCAAGATATGAGTACTTAAGTAAATGGG 5154
966 ySerAspAsnGlyProAlaPheValAlaGlnValSerGlnLysLeuAla 983
|||||
5155 GTCCGACAAATGAGACCGCTTGTGGCCAGTAAAGTACAGGATTTGGCA 5204
983 ySIIleuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 999
|||||
5205 CTCAGTTGGCATGATGGAATTAACATGTGCTTACCCGCCCAAGC 5254
1000 SerGlyGlnValGluArgMetAsnArgThrIleLysGluThrIleThrLy 1016
|||||
5255 TCAGGACAGGTAGAGAGATGAATAGAACATTAAGAGACCTTGATTA 5304
1016 sLeuThrThrGluThrGlyLysAsnAspTrpMetAlaLeuLeuProPhe 1033
|||||
5305 ATTGGCCATGAGACCGCGGGAAGACTGGGTGCTCTCTCCCTCC 5354
1033 aLleuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049
|||||
5355 CGCTTTCCGAGCCCGGACACCCCGAGAGTTTGGGCTCACCTCTTT 5404
1050 GluLeuLeuTyrGlyGlyProProProLeuAlaGlu.....IleAl 1063
|||||
5405 GAACTCTGATAGGGGAGCTCCCTTTAATAAAGATGTGGACATT 5454
1063 aPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgL 1080
|||||
5455 GTTCCCGATTCAGGCTCTGCTTACCTCTCT...TTCCTTATTCATT 5501
1080 euLysAlaLeuGluTyrValArgGlnArgAlaTrpLysGlnLeuArgGlu 1096
|||||
5502 TAAAGGCCCTGAAGATTAATAGACCCCAATTTGGGACCGCTAAAGACG 5551
1097 ALatYrSerGlyGlyAspLeuGlnValProHisArgPheGlnValGlyAs 1113
|||||
5552 GCCTACACCCAGGAGGACCGACCGAGTACCCACAGAGTCCAGGTCCGGGA 5601
1113 pSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpL 1130
|||||
5602 CCAAGTCTTGCTCAGACGACATCGAACCGGTAGCTTGAAACCAACGGTGA 5651
1130 ySGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGlu 1146
|||||
5652 AGGACCCCTTTAGTGTACTAACAACCTCTACGCGACGAGAAAGTTGAC 5701
1147 GlyIleProThrTrpIleHisAlaSerHisValLysProAlaProProP 1163
|||||
5702 GGGATTCCTCTCTGATCCACGCTCTCACGCTCAAGAGGCCCCCAGTCA 5751
1163 oAsp.....SerGlyTrpArgAlaGluLysThrGluAsn 1175
|||||
5752 AGATGAAGAAACCCAGCAAGCAACTGGGCTGTGGAAGCCATGATTAAC 5801
1175 rGluLeuArgLeuHisArgLeuValProTyrSerAsnAsnSer 1191
|||||
5802 CTCTTAAGCTTCTGCTGCGAGAGACCCCTC.....CATCAC 5842
1192 ProGly 1193
|||||
5843 CCTGGA 5848
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-716-351A-1
seq_documentation_block:
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Glibon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,351A
; FILING DATE:

```

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PC9/US94/03784
 FILING DATE: 06-APR-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Bastian, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 15280-128-1PC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 543-9600
 TELEFAX: (415) 543-5043
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8535 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1..8535
 OTHER INFORMATION: /standard_name="Galv SPATO Genome"
 US-08-716-351A-1

alignment_scores:
 Quality: 4319.00 Length: 1185
 Ratio: 4.238 Gaps: 6
 Percent Similarity: 85.992 Percent Identity: 68.523

alignment_block:
 US-09-171-553b-5 x US-08-716-351A-1 ..

Align seg 1/1 to: US-08-716-351A-1 from: 1 to: 8535

```

1  G1yArGArGlySerAspProLeuProGluProArgValThrLeuLysVa 17
   ||| .....|.....|.....|.....|.....|.....|.....|
2564 GGGAGTCAGGGGTTGGACCCCTCCCGAAGCTAGGTAACACTGACTGT 2613
   |||.....|.....|.....|.....|.....|.....|.....|
17  1GluGlyGlnProValGlnPheLeuValAspThrGlyAlaGlnHisSer 34
   ||||| .....|.....|.....|.....|.....|.....|.....|
2614 GGAGGGAGACCCCATGATGCTTCGTCGACACCGGAGCTGAACATTCAG 2663
   ||||| .....|.....|.....|.....|.....|.....|.....|
34  alleuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
   ||||| .....|.....|.....|.....|.....|.....|.....|
2664 TATTGACCCCAAGCCATGGGAAAGTAGGACGACGAGCGTCGTGGAA 2713
   ||||| .....|.....|.....|.....|.....|.....|.....|
51  G1yAlaThrGlyGlnArgGlnTyrProTrpThrArgArgThrValAs 67
   ||||| .....|.....|.....|.....|.....|.....|.....|
2714 GGAGCGACGAGCGACGACGACGACGACGACGACGACGACGACGACG 2763
   ||||| .....|.....|.....|.....|.....|.....|.....|
67  pleuGlyValGlyArgValThrHisSerPheLeuValIleProGluLys 84
   ||||| .....|.....|.....|.....|.....|.....|.....|
2764 AATTGGACATTAACAAAGTAGACCCATCTCTGCTGATATACCCGAGTGC 2813
   ||||| .....|.....|.....|.....|.....|.....|.....|
84  rovalProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
   ||||| .....|.....|.....|.....|.....|.....|.....|
2814 CTGCTCTCTCTTGGCGAGGACCTCTCAACCAACCAACCAACCAACCA 2863
   ||||| .....|.....|.....|.....|.....|.....|.....|
101 serPheGluGlnGlyArgProGluValSerValAsnLysProIleTh 117
   ||||| .....|.....|.....|.....|.....|.....|.....|
2864 CAGTTTCCTGCTGAGGCGCCACAGGTAACTAGGGAGAACCCCTACTAT 2913
   ||||| .....|.....|.....|.....|.....|.....|.....|
117 rValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGln 134
   ||||| .....|.....|.....|.....|.....|.....|.....|
2914 GTGCTGGTCTTAACCTGGAAGAAATACCGACTACATGAAAAAGCAGC 2963
   ||||| .....|.....|.....|.....|.....|.....|.....|
134 allYsProAspGlnAspIleGlnSerTrpLeuGlnGlnPheProGlnAla 150
   ||||| .....|.....|.....|.....|.....|.....|.....|
2964 TAA...CCCTCCTATTCAGCACCATCTGCTGCTGCTGCTGCTGCTG 3010
   ||||| .....|.....|.....|.....|.....|.....|.....|
151 TrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnVa 167

```

```

|||.....|.....|.....|.....|.....|.....|.....|
3011 TGGCGACAAAGAGCGCGCATGGAGCTAGCCAAATCAAGTCCACAGTGT 3060
   |||.....|.....|.....|.....|.....|.....|.....|
167  11leGlnLeuLysAlaSerAlaThrProValSerValArgGlnTyrPro 184
   ||||| .....|.....|.....|.....|.....|.....|.....|
3061 AGTAGAGCTAAGATCAGGTGCTTCCACAGGCGGTGCTGCTGCTGCTG 3110
   ||||| .....|.....|.....|.....|.....|.....|.....|
184  euserArgGluAlaArgGluGlyIleThrProHisValGlnArgLeuIle 200
   ||||| .....|.....|.....|.....|.....|.....|.....|
3111 TGACCAAGAAAGCTCGGAGAGGTATCAGACCCACATCCAGAAAGTCTTA 3160
   ||||| .....|.....|.....|.....|.....|.....|.....|
201  GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeu 217
   ||||| .....|.....|.....|.....|.....|.....|.....|
3161 GACCTAGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3210
   ||||| .....|.....|.....|.....|.....|.....|.....|
217  uProValArgGlySerProGlyThrAsnAspTyrArgProValGlnAspLeu 234
   ||||| .....|.....|.....|.....|.....|.....|.....|
3211 ACCTGTAAAAAGCCAGGAGCAATGACTATCGCCAGTTCAGACCTGA 3260
   ||||| .....|.....|.....|.....|.....|.....|.....|
234  rGluValAsnLysArgValGlnAspIleHisProThrValProAsnPro 250
   ||||| .....|.....|.....|.....|.....|.....|.....|
3261 GAGAAATTAATTAAGAGGTATACAGATATTCACACAGTCCAAACCT 3310
   ||||| .....|.....|.....|.....|.....|.....|.....|
251  TyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValle 267
   ||||| .....|.....|.....|.....|.....|.....|.....|
3311 TACAATCTTCTGAGTTCCTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 3360
   ||||| .....|.....|.....|.....|.....|.....|.....|
267  uAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnP 284
   ||||| .....|.....|.....|.....|.....|.....|.....|
3361 AGATCTCAAGATGCTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3410
   ||||| .....|.....|.....|.....|.....|.....|.....|
284  rLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
   ||||| .....|.....|.....|.....|.....|.....|.....|
3411 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3460
   ||||| .....|.....|.....|.....|.....|.....|.....|
301  LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePh 317
   ||||| .....|.....|.....|.....|.....|.....|.....|
3461 CTGACCTGAGCGGCTACACCAAGGTTCAAGAACTCTCCACCTCTCT 3510
   ||||| .....|.....|.....|.....|.....|.....|.....|
317  eAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProG 334
   ||||| .....|.....|.....|.....|.....|.....|.....|
3511 CGACGAGCGCTCCACCGAGATTTGCTTGGCGCCCTCAACCCCC 3560
   ||||| .....|.....|.....|.....|.....|.....|.....|
334  lValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThr 350
   ||||| .....|.....|.....|.....|.....|.....|.....|
3561 AGTGTGTTTACTCCAAATGTGAGACCTTGTGCTGCTGCTGCTGCTG 3610
   ||||| .....|.....|.....|.....|.....|.....|.....|
351  LysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGluLeuSerAs 367
   ||||| .....|.....|.....|.....|.....|.....|.....|
3611 TATGAAAGACTGCAAAAAGAACACAGACGCTTACAGAGTTAAGTAA 3660
   ||||| .....|.....|.....|.....|.....|.....|.....|
367  pleuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGlu 384
   ||||| .....|.....|.....|.....|.....|.....|.....|
3661 GTTGGGCTACCGGCTATTCGCTAAGAGGCGCCAGCTCTCCAGAGGAG 3710
   ||||| .....|.....|.....|.....|.....|.....|.....|
384  alThrTrpLeuGlyTyrSerLeuArgGlyGlnArgTrpLeuThrGlu 400
   ||||| .....|.....|.....|.....|.....|.....|.....|
3711 TCACCTATCTGGGGTCTACTACTACAGAAAGAAAGAAAGAGGCTAAC 3760
   ||||| .....|.....|.....|.....|.....|.....|.....|
401  AlaArgLysLysThrValAlaGlnIleProAlaProThrThrAlaLysG 417
   ||||| .....|.....|.....|.....|.....|.....|.....|
3761 GCCCGAAAGGCTACTCTTATGAATAATCCCTGCTTACAGCCGCCGAG 3810
   ||||| .....|.....|.....|.....|.....|.....|.....|
417  nValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProG 434
   ||||| .....|.....|.....|.....|.....|.....|.....|
3811 GGTCCGTGAATTCCTGAGCACTGCGGATTTCTGAGGCTGTGATCCCTG 3860
   ||||| .....|.....|.....|.....|.....|.....|.....|
434  lPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlyLysGly 450
   ||||| .....|.....|.....|.....|.....|.....|.....|
3861 GGTTCGTTCCCGGCTGACACCTTGTACCCCTTACCAAAAGAGACATC 3910
   ||||| .....|.....|.....|.....|.....|.....|.....|
451  GlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysL 467
   ||||| .....|.....|.....|.....|.....|.....|.....|

```



```

1066 sSerAlaSPValIleuLeuSerGlnProIleuPheSerArgLeuLysAlaL 1083
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5683 GGGTCCCGATAGATATTCCTCCCTGCTATTACTACTTAAGAGCTT 5732
1083 euGIuTrpValArgIleuAlaIleuArgIleuArgIleuArgIleuArgIleu 1099
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5733 TAGAAATGTAGAGAGCCCAATCTGGACCATCAAGAGGTGTAAAG 5782
1100 GtGtYAspLeuGlnValProHisArgPheGlnValGtYAspSerValTy 1116
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5783 CCGTGTACCTTAACATCCCTCAACCCGCTTCAGTGGGGATCAAGTGC 5832
1116 tValArgArgHisArgAlaGtYAspLeuGlnValArgIleuArgIleuArgIleu 1133
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5833 TGTCAGAGCCATCGACCCAGACCTTGGAGCTCGGTGGAAGAGCCAT 5882
1133 yIleuValIleuLeuThrThrProThrAlaValLysValGtYAspPro 1149
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5883 ACCTGGTGTGCTGACTACCCCGACCGCGGTAAAGTCGATGATTTGCT 5932
1150 ThrTrpIleuHisAlaSerHisValLysProAlaProPro.....ProAs 1164
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5933 GCCGGGTCCATGCTCTCACTCAACCAACGACACCTTCGGGACACAGA 5982
1164 pSerGtYTrpArgAlaGtYAspLeuGlnValProLeuLysLeuArgLeuH 1181
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5983 TGAGTCTCGGAGGAGCTGGAAAGACTGATCATCTTAACTGCGATATTC 6032
1181 tArG 1182
      :::::
6033 GGGCG 6037
seq_name: /cgn2_6/prodata/2/ina/5A.COMB.seq:US-08-258-420-13
seq_documentation block:
; Sequence 13, Application US/08258420
; Patent No. 5710037
; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gillfillian, Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,420
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-208
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; LENGTH: 8202 bases
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: viral genome
US-08-258-420-13

alignment_scores:
  Quality: 3993.50      Length: 1210
  Ratio: 3.982          Gaps: 11
  Percent Similarity: 82.893      Percent Identity: 62.727

alignment block:
US-09-171-553b-5 x US-08-258-420-13/rev ..

Align seg 1/1 to reverse of: US-08-258-420-13 from: 1 to: 8202

1 GtYArgArgGtYSerAspProLeuProGluProArgValThrLeuLysVal 17
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5529 GAGGTCAAGGTGAGGAGCCGCCCTCGAACCAGATTAACCTCAAGT 5480
17 tGtYGtYLeuProValGluPheLeuValAspThrGtYAlaGtYHisSerY 34
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5479 CGGGGGCAACCCGTCACCTTCCTGTAGATCTGGGGCAACCTCGG 5430
34 lleuLeuGlnProLeuGtYLeuLysGtYLysSerTrpValMet 50
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5429 TGCTGACCCCAAAATCTCGAACCCTTAAGATTAAGTCTCGTGGCCAA 5380
51 GtYAlaThrGtYLeuArgIleuTrpTrpThrThrArgArgThrValAs 67
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5379 GGGGCTACTGAGAGCAAGCGTATCGCTGACACCGATCGCAAGTACA 5330
67 pLeuGtYValGtYArgValThrHisSerPheLeuValIleuProGluCysP 84
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5329 TCTAGTACCGGTAAAGTCAACCTCTTCCCTCACTGACAGACTGTC 5280
84 roValProLeuLeuGtYArgAspLeuLeuThrLysMetGtYAlaGtYLeu 100
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5279 CTAATCTCTGTTAGAGAGATTTCTGACTTAACATAAAGCCCAATC 5230
101 SerPheGlu...GtYArgProGluValSerValAsnLysProIle 116
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5229 CACTTTGAGGATCAGAGCTCAGGTTATGAGCAATGGGAGGCCCTC 5180
116 eThrValLeuThrLeuGlnLeuAspGluTrpArgLeuTrpSerProG 133
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5179 GCAAGTGTGACCTTAATATAGAGATGAGTATCGGCTACATGACAGCT 5130
133 InValLysProAspGlnAspIleGlnSer...TrpLeuGlnGlnPhePro 148
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5129 CAAGAAGCCAGATGTTCTCTAGGGTCCACATAGGCTGTGATTTCT 5080
149 GtYAlaTrpAlaGtYThrAlaGtYMetGtYLeuAlaLysLysValProP 165
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5079 CAGGCTGGGGCGGAACCGGGGCGATGGACTGGCACTTGGCCAAAGCTC 5030
165 oGtYValIleGlnLeuLysAlaSerAlaThrProValSerValArgGln 182
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
5029 TCTGATCATACCTTGAAGCAACCTCTACCCCGTGCATTAACCAAT 4980
182 yTrpLeuSerArgGtYAlaArgGtYIleTrpProHisValGlnArg 198
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4979 ACCCATGTCAACAAGAGCCAGATGGGATCAAGCCCAACAACACAGA 4930
199 LeuIleGlnGtYLeuValProValGlnSerProTrpAsnTrp 215
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4929 CTGTTGACCGAGGCTGACTGTTACCTGCCAGTCCCTCGAAGACGCC 4880
215 oLeuLeuProValArgLysProGtYThrAsnAspTrpArgProValGln 232
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4879 CCTGCTACCGTTAAAGAACCGAGGACTAATGATTATAGCCGTGTCAG 4830

```

232 splenArgLysValAsnLysArgValGlnAspIleHisProThrValPro 248
|||||.....
4829 ATCTGAGAGAGTCAACAAGCGGTGAGACATCCACCCACCGTGGCC 4780
249 AsnProThrAsnLeuLeuSerAlaLeuProProGluAArgAsnTrpTrpTh 265
|||||.....
4779 AACCCCTTAACCTCTTGAGCGGGCTCCACCGTCCACACAGTGTACAC 4730
265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
|||||.....
4729 TGTGCTGATTTAAAGAGATGCTTTTCTGCTGAGACTCCACCCACCA 4680
282 eArgLysProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr 298
|||||.....
4679 GTACACCTCTCTCCCTTGAGTGAAGAGATCCAGAGATGGAAATCTCA 4630
299 GlyLysLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
|||||.....
4629 GGACAAATTGACCTGGACCAAGCTCCACAGGGTTTCAAAAACAGTCCAC 4580
315 rLLePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgLysGln 332
|||||.....
4579 CCTGTTGATGAGGACTGCACACAGACTTCGAGACTCCGATCCAGC 4530
332 iSPROGlnValThrLeuLeuGlnTrpValAspAspLeuLeuAlaGly 348
|||||.....
4529 ACCAGACTGATCCTGCTACAGTACGTGATGACTTACGTGCGCGCC 4480
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGlu 365
|||||.....
4479 ACTTCTGAGCTAGACTGCCAACAAGTACGTGCGCCCTGTTCAAAACCT 4430
365 uSerAspLeuGlyTrpArgAlaSerAlaLysAlaGlnIleCysArg 382
|||||.....
4429 AGGGAACCTCGGATCGGGCTCGGCCAAGAAAGCCCAATTGGCACA 4380
382 rGlnValThrTrpLeuGlyTrpSerLeuArgGlyGlnArgTrpLeu 398
|||||.....
4379 AACGGGCAAGTATCTGGGGTATCTTAAAGAGGGTACAGATGGCTG 4330
399 ThrGlnAlaArgLysLysThrValAlaGlnIleProAlaProThrAla 415
|||||.....
4329 ACTGAGGCCAAGAAAGACTGTGATGGCGGAGCCTTACCTCCGAAGCC 4280
415 aLysGlnValArgLysLeuLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
|||||.....
4279 TCACCAACTAAGAGAGTCTCTAGGAGCGAGCTCTGCGCTTGA 4230
432 lProGlyPheAlaThrLeuAlaAlaProLeuTrpProLeuThrLysGlu 448
|||||.....
4229 TCCCTGGGTTGCAAAATGGCAGCCCTTGTACCTCTCACCAAAACG 4180
449 LysGlyLysPheSerTrpAlaProGlnHisGlnLysAlaPheAspAla 465
|||||.....
4179 GGGACTCTGTTTAATGGGGCCCAAGCAACAAGAAAGCCCTATCAAGAA 4130
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
|||||.....
4129 CAAGCAAGCTCTTCAACTGCCAGCCCTGGGGTTGCCAGATTGACATA 4080
482 ySPROpHeThrLeuTrpValAspGluArgGlyValAlaArgLysVal 498
|||||.....
4079 AGCCCTTTAACTCTTTGTGACGAGAAAGCAAGGCTACGCCAAAGCGCTC 4030
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTrpLeuSerLys 515
|||||.....
4029 CTAAACGCAAAAGCTGGACCTGGCGTGGCGCGGCTTACCTGTCTTAA 3980
515 sLysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleA 532
|||||.....
3979 AAAGCTAGACCAAGTGGCAGCTGGCGCCCTGACCTACGATGCTGG 3930
532 laAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||.....
3929 CAGCATTTGCACTTCTGACAAAGATGCTGGCAAGCTCACTATGGACAG 3880
549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnP 565
|||||.....
3879 CCGTGTGCTATCTGGGCCCCCATGCTGCTAGAGCACTAGTAAAGCAAC 3830
565 oProAspArgTrpPheThrAsnAlaArgMetThrHisTyGlnSerLeu 582
|||||.....
3829 CCCTGATGCTGGCTGCTCAATGCTCCGAGATGCCATTTACCAAGCCCTG 3780
582 eLeu...ThrGluArgValThrPheAlaProProAlaLeuAsnPro 597
|||||.....
3779 TCTGAGACAGGACCGGGTCCAGTCTGGGCTAGTGGCCCTTAAATCCA 3730
598 AlaThrLeuLeuProGlnGluTrpAspGluProValThrHisAspLysH 614
|||||.....
3729 GCTACGCTGCTCCT...CTGCTGAGAGAGGGCTCCAACTGACTGCT 3683
614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
|||||.....
3682 TGACATCTTGCTGCTGAGCCACGACCTAGATCATCTAGGACCCAC 3633
631 rLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTrpVal 647
|||||.....
3632 CCTCCCAAGCCGACACACCTGCTGACAGATGGAGCAGCTTCG 3583
648 ValGluGlyLysArgMetAlaGlyAlaValAlaValAspGlyThrArgTh 664
|||||.....
3582 CAAGAAGGCGACGCTAAGCGCGGAGCAGCGGTGACCTAGACATGAGGT 3533
664 rIleThrAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 681
|||||.....
3532 AATCTGGGCGAGGATTCGACGCGGACATCGGCCCAAGAGCTGAC 3483
681 eMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
|||||.....
3482 TATAGCGCTCACCCAGCGCTTAAAGATGGCAGAAAGTAAAGCTTAAT 3433
698 lLeuTrpThrAspSerArgTrpAlaPheAlaThrAlaHisValHisGlyVal 714
|||||.....
3432 GTTATACTGATAGCGCTTACGCTTTTCCACCGCCCTATTCATGAGCA 3383
714 aLleuTrpGlnAlaGlyLeuLeuThrSerAlaGlyArgLysVal 731
|||||.....
3382 AATATTAAGAAAGCGCGGGTCTCACAATCAGAAAGAAAGATCAAGA 3333
731 sLysGluGluIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArg 747
|||||.....
3332 ACAAGAGAGAGATCTTACCTCTTAAAGGCTCTCTTGTGCCCAAAAGA 3283
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleS 764
|||||.....
3282 CTTAGCATTAATTCATTCGCCGCGGACATCAAAAAGAAAGACGCGAGGC 3233
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGln... 779
3232 CAGGGGCAACCGGATGGCGGCAAGCGCCGAGAGTGGCCACTAGAG 3183
780AlaValAsnLeuLeuProIleIleGluTrpProLysAla 792
3182 AAATCCAGAGACTTCACACTTGTGATACAAAATCAACCCCTATAC 3133
793 ProGluProArgArgLysTrpThrLeuGluAspTrpGlnIleLysLys 809
|||||.....
3132 CATGAA...CACTTTCATTAACAGTACAGACAAAGATTTGACCA 3086
809 sLLeAspGlnPheSerGluThrProGluGlyThrCysTrpTrpSer... 824
3085 ACTA.....GGAGCCACTTATGACAGCGCA 3060
825TrpGlyLysGluIleLeuProHisLysGluGly 835
|||||.....

```

3059 AGAATATTGGCTCAGTCAAGAAAGCCTGTTATGCTGATCAATTCACC 3010
836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrHis 852
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3009 TTGAGTACTAGAGATCTTCTACCAATTCAGCCACTCAGCTTCCTCAA 2960
852 sLeuGlnIleValArg.....ThrSerProTyrHisValLeuArgL 867
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2959 AACCAAGGCTCTCTAGAGAGAACCCCAAGCTCTACTACTAGTGTGAAC 2910
867 euProGluValAlaAspSerValValLysHisCysValProCysGlnLeu 883
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2909 GGGATCGAAGACTCAAAATATTCATCTGAGCTGACCAAAAGCTTGACAA 2860
884 ValAsnAlaAsnProSerArgIleProProGlyTyrArgLeuArgGlyse 900
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2859 GTCAATGCCAGCAAGTCTGCCGTTAAGCAAGAACTAGGTCGCCGGGCA 2810
900 rHisProGlyAlaHisTrpGluValAspPheThrGluValProAlaL 917
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2809 TCGGCTCGTGCACACACTGGAGATGATGATTCACGAGGTAAACCTGGAT 2760
917 yTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTyr 933
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2759 TGTATGGCTATAGATCTTTAGTTTGTAGATACTTTTCTGGCTGG 2710
934 ValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 950
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2709 ATAGAGAGCTTCCCAACTAAGAAAGAACCCGCAAGTCGTCGACCAAAA 2660
950 sIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlyS 967
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2659 ACTGCTAGAGAGATCTTCCAGTTCGGCATGCCGACGAGTATGGGAA 2610
967 eAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2609 CTGACAAATGGCGCTGCTCTCCCAAGGTAGTCAGACAGTGGCCGAT 2560
984 IleLeuGlyIleAspTyrLysLeuHisCysAlaTyrArgProGlnSerSe 1000
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2559 CTGTGGGGATGTGATTGAAATTACATGTGCATACAGACCCCAAGCTC 2510
1000 rGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2509 AGGTCAAGTAGAAGAAATAGATAGACACCTCAAGAGATTTAACTAAT 2460
1017 euThrThrGluThrGlyIleAsnAspTyrMetAlaLeuLeuProPheVal 1033
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2459 TTAAGCTTGCACACTGCTCTAGAGACTGGTCTCTACTCTCCCTTAGCC 2410
1034 LeuPheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProTyrG 1050
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2409 CTGTACCGAGCCCGCAACAGCGCGGCCCCCACTGAGCTTCAACCCATAT 2360
1050 uLeuLeuTyrGlyGlyProProProLeuAlaGlnIleAlaPheAlaHis 1067
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2359 GATCTTATATGAGGACACCCCGCCCTTGAATCTCCCTGACCCCGACA 2310
2359 eAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2309 TGACCAAGATTACTAACAGCCCTCTCTCCAAAGCTCACTTACAGGCCCT 2260
1084 GluTyrValArgGlnArgAlaTyrLysGlnLeuArgGluAlaTyrSerG 1100
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2259 CAAGCACTACACAGAGGTCTGGAAAGCCACTGCTGCGGCTCATATAGA 2210
1100 yGlyAspLeuGln.....ValProHisArgPheGlnValGlyAspSerV 1115
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2209 CCAGCAAGACACGACAGTGTATACACACCCCTTCGCTGTCGCGACACG 2160
1115 alTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTyrLysGly 1131
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2159 TGTGGTATACCGCGGACCAAGACTTAAGAACTTGGAACTCTGTTGAAAGA 2110

```

```

1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluLyl 1148
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2109 CCTATACCTCTCTGCTGACACACCCACCCGCTCTCAAGTATAGCGCAT 2060
1148 eProThrTrpIleHisAlaSerHisValLysProAlaProProProAsp 1165
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2059 CGCTGGTGTATCAGCGCCGCTCCTCAAGGCGGACAAACCCCTCCG 2010
1165 eArgL.....TrpArgAlaGluLysThrGluAsnPro 1175
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2009 CCGAAGACAGATCAGACGACGACATGGAGGTCCAGCGTTCTCAAAACCC 1960
1176 LeuLysLeuArgLeuHisArgLeuValPro 1185
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1959 TTAAAGTATAGATTAAACCCGTGGGGCCCC 1930

```

seq_name: /cqn2_6/plodata/2/ina/5A_COMB.seq:US-08-110-300A-8

seq_documentation_block:

Sequence 8, Application US/08110300A

Patent No. 5643756

GENERAL INFORMATION:

APPLICANT: Pinter, Abraham

APPLICANT: Kayman, Samuel

TITLE OF INVENTION: FUSION GLYCOPROTEINS

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Davis Hoxie Faithfull and Hapgood

STREET: 45 Rockefeller Pl.

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/110,300A

FILING DATE: 20-AUG-1993

CLASSIFICATION: 435

AUTHOR/AGENT INFORMATION:

NAME: Jacobs, Seth H.

REGISTRATION NUMBER: 32,140

REFERENCE/DOCKET NUMBER: 11698A50

TELEPHONE: 212-757-2200

TELEFAX: 212-586-1461

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 8323 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-110-300A-8

alignment_scores:

Quality: 3980.00

Ratio: 3.992

Percent Similarity: 82.670

Percent Identity: 62.687

alignment_block:

US-09-171-553B-5 x US-08-110-300A-8 ..

Align seg 1/1 to: US-08-110-300A-8 from: 1 to: 8323

1 G1ATGATGAGGlySerAspProLeuProGluProArgValThrLeuLysVa 17

614 sGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleP 631
:::|||||
4083 TGACATCTTGCTGAGCCACGGAACCTAGACCAAGATCTACGGACGAC 4132
631 rGluThrGlyGluValLeuThrTrpPheThrAspGlySerSerGlyVal 647
|||||
4133 CTCCTCCAGACGCTGACACACTGGTGGTACACAGATGGAGCACTCTCTG 4182
648 ValGluGlyLysArgMetAlaGlyAlaValAlaValAspGlyThrArgTh 664
|||||
4183 CAAAGAGGGGCAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4232
664 rIleThrAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 681
:::|||||
4233 AGCTGGGCGCAAGCACTGCGGCGGAGCATCGGCGCAAGAGCTGAGT 4282
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
|||||
4283 TGATAGCGCTCACCCAGGCTTAAATGGCAAGAGTAAAGAGCTGAGT 4332
698 IleThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
:::|||||
4333 GTTACACCGATAGCGCGTATGCTTTGGCACTGCCATATTCACGAGAGA 4382
714 aIleThrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
|||||
4383 AATATATAGAGGCGCGGGTCTTCACATCAGAGAGGAGGAGGAGGAGGAG 4432
731 sLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
|||||
4433 ATAGAGACAGATCTTGCGCTTACGAGAGCTCTCTCTGCGGCAAGAGA 4482
748 LeuAlaIleIleHisCysProGlnHisGlnLysAlaLysAspLeuIleSe 764
:::|||||
4483 CTAGCATATTCATTCGCCGAGACATCAGAGGAGGAGGAGGAGGAGGAG 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
|||||
4533 AAGGGGCAACAGAGATGGCCGACCAAGCGCGGAGAGTAAAGTAAAGTAA 4581
781 aAlaLeuLeuProIleIleGluThrProLys..... 791
4582GAAACTCCAGAGACTTCCACTTCGATA 4611
792AlaProGluProArgArg.....GlnTyrThrLeuGluAs 803
|||||
4612 GAAATTCAGCCCCCTATCTACTCATGAACATTTTCACTATACGATGACTGA 4661
803 pTrpGlnGluIleLysLysIleAspIleHisSerGluThrProGluGlyT 820
:::|||||
4662 CATTAAGATCTGACTTAAGTGGGCCACTATGACGATGCAAA...A 4708
820 hrcysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
|||||
4709 AGCTGTGGGTTTATCAGGAGAACCTGTATGCTGATCAATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
|||||
4759 GAACATATTAGATTCTTCTCATCAATGACCACTCAGCTTCATCAAAAAC 4808
853 uGlnGlnLeuValArgThrSer.....ProThrHisValLeuArgLeuP 868
:::|||||
4809 AAAGGCTCTTGTAGAAAGACTACTGCTCTTATTTACATGCTGACCGGG 4858
868 rGluValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
:::|||||
4859 ATGCAAGCGTCAAAAGACATCATGAGACTGCAAGGCTGTGACAGGTC 4908
885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerHis 901
|||||
4909 AATGCCAGCAAGTCTGCCGTCAAAACAGAGGAGTGAAGTTGAGAGGACCG 4958

901 sproGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
:::|||||
4959 ACCCGGACCCACTGGGAATGTATGTTCACTGAGTAAACCTGGCCTGT 5008
918 yrcGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTyrPval 934
|||||
5009 ATGGGTAAATATATCTTTAGTTTCATAGACACTTCTCTGTGATGGGTA 5058
935 GluAlaLysProThrLysGlnLeuThrSerThrValValAlaLysLysI 951
|||||
5059 GAACGTCTCCCAACCAAGAAAGAACTGCCAAAGTTGTAAACCAAGAACT 5108
951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
:::|||||
5109 ACTAGAAAGAAATCTTCCCGAGATTTGGCATGGCCACAGATTTGGAAACG 5158
968 sPasnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
:::|||||
5159 ACAATGGGCGCTGCTGCTCCAAAGTAAAGTACAGACAGATGCGGATTTA 5208
985 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerG 1001
|||||
5209 CTGGGGGTGATTTGAAACTACATTTGCTTACAGACCCACAGATTCAG 5258
1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
|||||
5259 TCAGGTGAAAGAAATGATAGGACATCAAGAGACTTAACTTAAATTTGA 5308
1018 hThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
:::|||||
5309 CGCTTGCAACTGCTGAGGAGCTGGGTGCTGCTGCTGCTGCTGCTGCTG 5358
1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
:::|||||
5359 TATGAGCGCCGACACAGCCGCGGCGCCAGTGTCTCACCACCATTTGAAT 5408
1051 uLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerA 1068
:::|||||
5409 CTATATGCGGCGACCCCGCCCTTGTAAACTTCCGATCCGATGACATGG 5458
1068 lAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
:::|||||
5459 CAAGGTATTCATCAATACCCCTCTCCAAAGCCCATTTTACAGGACTCTAC 5508
1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGlyI 1101
:::|||||
5509 CTGGTCCGACGAGAACTGTGAGACCCGTTGGCGGAGTTTACAA...GA 5555
1101 yAspLeuGln.....ValProHisArgPheGlnValAlaLysSerV 1115
:::|||||
5556 ACAACTGGACGCGCGGTAGTGGCTTCCGATCCGATGCGTACACAG 5605
1115 aTyrValAlaArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
|||||
5606 TGTGGTCCGAGACACCAACTAAATAATCTAGAAACCCCGCTGAAAGGA 5655
1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyI 1148
|||||
5656 CTTTATACCGCTCTCTACTAGTACCCGCGGTCTCAAAAGTGGAGCGCAT 5705
1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
:::|||||
5706 TGCAGCGTGGATCCAGCTGCGCCAGTAAAGGCTCCGACACACAGATTTG 5755
1161 ..ProProAspSerGlyTrpArgAlaGluLysThrGlnAsnProLeu 1176
:::|||||
5756 AACCAACATCGGAGATCGACATGGCGTTTCAACGCTTCAAAATGCCCTA 5805
1177 LysLeuArgLeuHisArg 1182
|||||
5806 AAGATTAAGATTGACCCCG 5823


```

seq_documentation_block:
; Sequence 8, Application US/08886642
; Patent No. 5952474
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 45 Rockefeller Plaza, Suite 2800
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/886,642
; FILING DATE: 01-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/110,300
; FILING DATE: 20-AUG-1993
; APPLICATION NUMBER: 07/938,100
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hone, William J.
; REGISTRATION NUMBER: 26,739
; REFERENCE/DOCKET NUMBER: 07/63/010002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212/765-5070
; TELEFAX: 212/258-2291
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-886-642-8

alignment_scores:
      quality: 3980.00      length: 1206
      ratio: 3.992         gaps: 12
      percent similarity: 82.670      percent identity: 62.687

alignment_block:
US-09-171-553b-5 x US-08-886-642-8
Align seq 1/1 to: US-08-886-642-8 from: 1 to: 8323

1  G1YARgArGgYserAsPrProLeuPProG1uPrArGValThrLeuTysVa 17
   |||.....|||.....|||.....|||.....|||.....|||.....
2236  GGAAGTCAAGGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2285
   |||.....|||.....|||.....|||.....|||.....|||.....
17  1G1uG1G1AProVa1G1uPheLeuValAsPThrG1A1aG1uH1sery 34
   |||.....|||.....|||.....|||.....|||.....|||.....
2286  CGGGGGGCAACCCCTCACCTTCTAGTGAATGAGGAGAGAGAGAGAGAG 2335
   |||.....|||.....|||.....|||.....|||.....|||.....
34  aLeuLeuGlnProLeuGlyLysLeuLysG1uLysLysSerTrpValMet 50
   |||.....|||.....|||.....|||.....|||.....|||.....
2236  TGCTGACCAAAATCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2285
   |||.....|||.....|||.....|||.....|||.....|||.....
51  G1YAlaThrG1uGlnArG1uYrPrOTrPThrThrArgArGThrValas 67
   |||.....|||.....|||.....|||.....|||.....|||.....
2286  GGGGCTACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435
   |||.....|||.....|||.....|||.....|||.....|||.....
67  pLeuGlyValG1YArGValThrHisSerPheLeuValIleProGluCysP 84

```

```

|||||.....|||||.....|||||.....|||||.....|||||.....
2436  CTAAGCAGCCGAGTAAAGTCAACCATTTCTTCTTCTTCTTCTTCTTCTT 2485
   |||.....|||.....|||.....|||.....|||.....|||.....
84  rovalProLeuLeuGlyArGAsPLeuLeuThrLysMetG1A1aG1u 100
   |||.....|||.....|||.....|||.....|||.....|||.....
2486  CCTATCCTCTGCTAGAGAGAGATTTGCTGACTAAACTAAAGAGCCCAAT 2535
   |||.....|||.....|||.....|||.....|||.....|||.....
101  SerPheGlu...GlnGlyArGProG1uValSerValasMetProl 116
   |||.....|||.....|||.....|||.....|||.....|||.....
2536  CACTTGAAGGATCAGAGAGCTCAGGTTGTGGAGCAATGGAGAGAGCCCT 2585
   |||.....|||.....|||.....|||.....|||.....|||.....
116  eThrValLeuThrLeuGlnLeuLysAsPGLuYrArGLeuYrSerPro 133
   |||.....|||.....|||.....|||.....|||.....|||.....
2586  GCAAGTGTGACCTTAACATAGAGATGAGTATCGGCTACATAGAGAGCT 2635
   |||.....|||.....|||.....|||.....|||.....|||.....
133  lNValYrProAsPGLAsP1leGlnSer...TrPLeuGlnPhePro 148
   |||.....|||.....|||.....|||.....|||.....|||.....
2636  CAAAGGGCCAGATGCTCTTGAAGGTCACATGGCTCTCTGATTTCCTCC 2685
   |||.....|||.....|||.....|||.....|||.....|||.....
149  GlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValPro 165
   |||.....|||.....|||.....|||.....|||.....|||.....
2686  CAGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 2735
   |||.....|||.....|||.....|||.....|||.....|||.....
165  oGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgLnt 182
   |||.....|||.....|||.....|||.....|||.....|||.....
2736  TGTGATCATACCTCTGAAGGCAACCTTACCCCGTGTCCATTAACAT 2785
   |||.....|||.....|||.....|||.....|||.....|||.....
182  YrProLeuSerArGGLuAlaArgGluGlyLettProHisValGlnArg 198
   |||.....|||.....|||.....|||.....|||.....|||.....
2786  ACCCATGTCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2835
   |||.....|||.....|||.....|||.....|||.....|||.....
199  LeuIleGlnGlnGlyLeuValProValGlnSerProTrpAsnThrPr 215
   |||.....|||.....|||.....|||.....|||.....|||.....
2836  CTGCTGATCAGAGAGATTTCTGTAACCTTGCAGATCCCTTGAACAGCC 2885
   |||.....|||.....|||.....|||.....|||.....|||.....
215  oLeuLeuProValArG1uYrProG1YThrasAsPTrArGProValGln 232
   |||.....|||.....|||.....|||.....|||.....|||.....
2886  CCTGCTACCCGTTAAGAAACCGGGGAGTATGATTTAGGCTGTCCAG 2935
   |||.....|||.....|||.....|||.....|||.....|||.....
232  sPLeuArGGLuValAsnLysArGValGlnAsP1leHisProThrValPro 248
   |||.....|||.....|||.....|||.....|||.....|||.....
2936  ATCTGAGAGAGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2985
   |||.....|||.....|||.....|||.....|||.....|||.....
249  AsnProTyAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyTh 265
   |||.....|||.....|||.....|||.....|||.....|||.....
2986  AACCTTACACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3035
   |||.....|||.....|||.....|||.....|||.....|||.....
265  rValLeuAsPLeuLysAsPAlaPhePheCysLeuArgLeuHisProTh 282
   |||.....|||.....|||.....|||.....|||.....|||.....
3036  TGTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3085
   |||.....|||.....|||.....|||.....|||.....|||.....
282  eGlnProLeuPheAlaPheGluTrpArgAsPProG1YrGlnArgThr 298
   |||.....|||.....|||.....|||.....|||.....|||.....
3086  GTACATCTCTCTTCCCTTGTGAGAGAGAGATCCAGAGATGGAGATCCA 3135
   |||.....|||.....|||.....|||.....|||.....|||.....
299  G1YGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProTh 315
   |||.....|||.....|||.....|||.....|||.....|||.....
3136  GAGCAATTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3185
   |||.....|||.....|||.....|||.....|||.....|||.....
315  rIlePheAsPGLuAlaLeuHisArgAsPLeuAlaAsnPheArg1leGln 332
   |||.....|||.....|||.....|||.....|||.....|||.....
3186  CCTGTTGATGAAGCCCTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 3235
   |||.....|||.....|||.....|||.....|||.....|||.....
332  sPProGlnValThrLeuLeuGlnTrpValAsPAsPLeuLeuAlaGly 348
   |||.....|||.....|||.....|||.....|||.....|||.....
3236  ACCCAGACCTGATTTGCTCCAGTATGATGAGACTTCTCTGAGCCGCG 3285
   |||.....|||.....|||.....|||.....|||.....|||.....
349  AlaThrLysGlnAsPcysLeuGlnGlyThrLysAlaLeuLeuGluLe 365
   |||.....|||.....|||.....|||.....|||.....|||.....
3286  ACTTCTGAGCTTGAAGTCAACAGATGAGAGAGAGAGAGAGAGAGAG 3335
   |||.....|||.....|||.....|||.....|||.....|||.....
365  uSerAsPLeuGlyTrArGAlaSerAlaLysLysAlaGlnIleCysArg 382
   |||.....|||.....|||.....|||.....|||.....|||.....

```

```

3336 AGGGACCTCGGATATCGGCGCTCGCCAGAAAAGCCAAATTTCACAGA 3385
3382 rgluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeu 398
3386 AACAGGTCAAGTATCTGGGCTATCTCTTAAAGAGGTCACAGATGGGTG 3435
3399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
3436 ACTGAGGCCAGAAAAGAGACTGTGATGGGACGCTTACCTCCGAAAGACCCC 3485
415 alysluValAlaArgLuphLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
3486 TCGAACACTAAGGAGGAGTCTAGGAGCGAGGCTTGTGCGCTCTGGA 3535
432 lProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
3536 TCCCTGGGTTTCAGAAATGCGAGCCCTTGTACCTCTGCACCAAAAGC 3585
449 LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaI 465
3586 GGGACTCTGTGTGAGTGGGGCCAGACCAAGAAAGGCTTACCAAGAGAT 3635
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
3636 CAAGCGGCTCTTAACTGCCCCGCTGCGGATTCACAGCTTGACTA 3685
482 ySProPheThrLeuTyrValAlaSpGluArgLysGlyValAlaArgGlyAl 498
3686 AGCCCTTCAGACTTTTGTGACGAGAAACAGGCGTACGCCAAAGGTGTC 3735
499 LeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerL 515
3736 CTAAACCAAAACTGGGCGCTTGGCGCGCGGCTGCTTACCTGTCCAA 3785
515 sLysLeuAspProValAlaSerGlyTyrProValCysLeuLysAlaIleA 532
3786 AAAGCTAGACCCAGTGGAGCTGGGCGCCCTTGTCTACGAGATGGTACG 3835
532 lAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
3836 CAGCCATCGCGCTTGTACCAAGAGCGTGGCAAGCTCACCATGGGACAG 3885
549 AsnIleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnTr 565
3886 CCACTACTACTTCTGCCCCCATGCAGTAGAGGCACTACTTAGCAACACC 3935
565 oProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3936 CCCTGATCGCTGGCTCCAAAGCCCGAATGACCCCACTACCAAGCTCTGC 3985
582 eLeu...ThrGluArgValThrPheAlaProProAlaAlaLeuAsnPro 597
3986 TTCTGGACAGGAGCGAGTCCAGTTCGACCAATAGTGGCCCTTAAACCCA 4035
598 AlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysH 614
4036 GCTACGCTGCTCCT...CTACCTGAGGAGGCGGCTGCACATGACTGCTC 4082
614 sGlnLeuLeuIleGluLeuThrGlyValAlaArgLysAspLeuThrAspIle 631
4083 TGACATCTTGGCTGAAGCCCAAGCACTAGACCGATCTTACGACAGCAGC 4132
631 rLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTrpVal 647
4133 CTCTCCACAGCGCTGACCAACCTGTGTACACAGATGGAGAGCAGCTTCTG 4182
648 ValGluGlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgTh 664
4183 CAAGGAGGCGAGCGCAAGGCGGAGCAGACTAACCAACCGAGACGAGGT 4232
664 rIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 681
4233 AGCTTGGGCAAAAGCATGTCAGCGCGGACATCGGCCCAAGAGAGCTGAGT 4282
681 euMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsn 697
4283 TGATAGCGCTACCCAGAGCCTTAAATAATGCGACAAAGTAGAAGCTGAAT 4332
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
4333 GTTTACACCGATAGCCCTTATGCTTTGGCCACTGCCCATATTCCAGCAGAG 4382
714 aIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
4383 AATATATAGAAAGCGCGGTTGCTCACATPAGAAAGAAAGAAATCAAAA 4432
731 snLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArg 747
4433 ATAAGCAGAGATCTTGCCCTTACTGAAGGCTCTCTTCGCCCCAAAGAGA 4482
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleS 764
4483 CTTAGCATTAATTCATTGCCCGGACATCAGAAAGGAAACCGCGGAGGC 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaV 781
4533 AAGGGCAACAGAGATGGCGCACCAAGCGGCCGAGAACTAGCCACTAGA. 4581
781 aLAsnLeuLeuProIleIleGluThrProLys..... 791
4582 .....GAAACTCCAGAGACTTCCACACTTTCGATA 4611
792 .....AlaProGluProArgArg.....GlnTyrThrLeuGluAs 803
4612 GAAATTCAGCCCTTACTACTACATGACATTTTGCATATAGGAGTCTGA 4661
803 pTrpGlnIleLysLysLysIleAspGlnPheSerGlnThrProGluGlyT 820
4662 CATAAAGATCTGACTTAAGTAGGGCCACTTATGACATGCAAG...A 4708
820 hCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
4709 AGTGTGGGTATTCAGGAAAGCCTGTAAAGGCTGATCAATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrHisLys 853
4759 GAACATTAATGATTTCTCATATTAATTAACCCACTGACTTCTCAAAAC 4808
853 uGlnGlnLeuValArgThrSer.....ProTyrHisValLeuArgLeuP 868
4809 AAAGCTCTTTCAGAAAGAACTACTGTCTTATTAATGTCGAGACCGGG 4858
868 rGluValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
4859 ATCGAAGCGCTCAAGACATCACTGAGACTTGGCAAGCCGTGTCACAGGTC 4908
885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySerH 901
4909 AATGCGAGAGTCTGCGCTCAACAAAGGAGCTAGATTCGAGGGCAGCG 4958
901 sProGlyAlaHisTrpGluValAlaSpPheThrGluValLysProAlaLysT 918
4959 ACCGGCACCACTGGGAATGATTTCACTGAGGTAACCACTGGCGCTGT 5008
918 yGlyAsnLysTyrLeuLeuValAlaPheValAspThrPheSerGlyTrpVal 934
5009 ATGGGTATTAATCTTTAGTTTCAATAGACACTTCTGTGATGGGTA 5058
935 GluAlaTyrProThrLysLysGluThrSerThrValAlaAlaLysLysI 951
5059 GAAGCTTTCCCAACCAAGAAAGAACTGCCAAAGTTGTAACCAAGAGCT 5108
951 eLeuGluGluIlePheProArgPheGlyLysIleProLysValIleGlySerA 968
5109 ACTAGAAAGAAATCTTCCAGATTCGCAATGCGACAGGATATTGGGAACCG 5158

```

```

;
;      ZLP:  10111
;
;      COMPUTER READABLE FORM
;

```

116 ethrValLeuthrLeuGlnLeuaspaspGluTyrargLeuTyrSerf

2586 GCAAGTCTGACCTAAACATAGAAGATGAGTATCGGCTACATGAGACCT 2635
 133 InVallyProAspGlnAspIleGlnSer...TrrleuGlnGlnPhePro 148
 2636 CAAAAGGGCCAGATGTGCTTACAGGCTCAGATGGCTCTCTGATTTTCCC 2685
 149 GlnAlaTrpAlaGlnThrAlaGlyMetGlyLeuAlaLysGlnValaProPr 165
 2686 CAGGCGCTGGCCAGAAACCGGGGCTGGGCTGGCCGCTGCCACAGCTCC 2735
 165 ocLInValIleGlnLeuLysAlaSerAlaThrProValSerValArgGlnTr 182
 2736 TCTGATCATCTCTGAAAGGACACCTACCCCGCTGCCATAAACAAAT 2785
 182 yrrProLeuSerArgGlnAlaArgGlyIleTrrProHisValGlnArg 198
 2786 ACCCATGTCCAGAGAGCCAGACTGGGGATCAAGCCCCACATACAGAGA 2835
 199 LeuIleGlnGlnGlyIleLeuValProValGlnSerProTrrPasnThrPr 215
 2836 CTGCTGGATCAGGGAATTCGTGTACCTGCCAGCTCCCTGCCAACAGCC 2885
 215 oleuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlna 232
 2886 CCGCTACCCGTTAAGAAACCGGGGACTAATGATTATAGGCTGTCCAGG 2935
 232 spleuArgGlnValaAsnLysArgValGlnAspIleHisProThrValPro 248
 2936 ATCTGAGAGAGTCAACAGCGGTGGAGAGATCCACCCACCCCTGTGCC 2985
 249 AsnProTyrAsnLeuSerAlaLeuProProGlnArgAsnTrrTyrTh 265
 2986 AACCTTACACCTCTTGAGCGGGCTCCACCCGCTCCACACATGGTTACAC 3035
 265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
 3036 TGTGCTTGACTTAAAGATCTTTTTCCTGAGACTCCACCCACACA 3085
 282 erGlnProLeuPheAlaPheGlnTrrArgAspProGlyThrGlyArgThr 298
 3086 GTCTAGTCTCTTCGCTTGAAGTGAAGATCCAGAGATGGGAATCTCA 3135
 299 GLyGlnLeuThrTrrArgLeuProGlnGlyPheLysAsnSerProTh 315
 3136 GGACAATTTAACTGTGACCACTCCGCAAGGTTTCAAAAACAGTCCAC 3185
 315 rIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnH 332
 3186 CCTGTTGATGAAGCCTGCACAGGACTCGCAGACTTCGGGATCCAGC 3235
 332 IsrProGlnValThrLeuLeuGlnTrrValAlaAspAspLeuLeuAlaGly 348
 3236 ACCCAGACCTGATTTCTCCAGATATGATGACTTACTGCTGCCGCC 3285
 349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuGluLe 365
 3286 ACTTGTGAGCTTGACTGTCAACAGAGTACCGGGCCCTGTTACAACACCT 3335
 365 userAspLeuGlyTrrArgAlaSerAlaLysLysAlaGlnIleCysArgA 382
 3336 AGGGGACCTGGATTCGGGCTCGGCCAGAAAGCCCAATTGTCGAGA 3385
 382 rGlnValaThrTrrLeuGlyTrrSerLeuArgGlyGlyGlnArgTrrLeu 398
 3386 AACAGGTCAAGTATCTGGGGTATCTTCAAAAGAGGGTCAAGATGGCTG 3435
 399 ThrGlnAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
 3436 ACTGAGGCCAGAAAGAGAGCTGTGATGGGCAAGCCTACTCCAGAGACCCC 3485
 415 aLysGlnValaArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
 3486 TCGACAACTTAAGGAGTTCCTAGGAGCGCAGGCTTCTGTGCTCTGGA 3535
 4432 IeProGlyPheAlaThrLeuAlaAlaLarProLeuTrrProLeuThrLysGlu 448
 3536 TCCCTGGGTGTTCAGAAATGGCACCCCTTGTATCCCTGTCCACCAAGC 3585
 449 LysGlyGlyPheSerTrrPalaProGlnHisGlnLysAlaPheAspAlaI 465
 3586 GGGACTGTGTGTGAGTGGGGCCAGACAGCAAGAAAGGCTTACCAAGAT 3635
 465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValaThr 482
 3636 CAGACAGGCTCTCTTAAGTCCCTGCCCTGGATTCAGACTGTGACTA 3685
 482 ysrProPheThrLeuTrrValaAspGlnArgGlyValaIleArgGlyVal 498
 3686 ACCCTTGCAACTTGTGTGACGAGAGCGGGGTACGCCAAAGAGTGTG 3735
 499 LeuThrGlnThrLeuGlyProTrrArgArgProValAlaTrrLeuSerLys 515
 3736 CTACCGCAAAACCTGGGCTTGGCGCTGGCGGTGGCTACCTGTCCAA 3785
 515 sLysLeuAspProValAlaSerGlyTrrProValCysLeuLysAlaIleA 532
 3786 AAGCTAAGACCCAGTGGGAGCTGGGTGGCCCTGTGCTACGAGTGTAG 3835
 532 IaAlaValaAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
 3836 CAGCCATCCCGCTTGTGACGAAAGCGGTGGCAACCTCACCTGAGGACG 3885
 549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPr 565
 3886 CCACATGATCTTTGTGGCCCCCATGAGTCAAGGAGGACATGTTAAACAC 3935
 565 oProAspArgTrrPmetThrAsnAlaArgMetThrHisTrrGlnSerLeuL 582
 3936 CCTGATCGGTGCTCTCCACACCCCGAATGACCACTACCGAGCTGTGC 3985
 582 euLeu...ThrGluArgValaThrPheAlaProProAlaAlaLeuAsnPro 597
 3986 TTTGGACAGGAGCGAGTCCAGTTCGGACCAATAGTGGCCCTAAACCA 4035
 598 AlaThrLeuLeuProGlnGlnThrAspGlnProValaThrHisAspCysHis 614
 4036 GCTACGCTGTCTCCT...CTACCTGAGAGAGGGCTGCAACATGACCTG 4082
 614 sGlnLeuLeuIleGlnGlnThrGlyAlaArgLysAspLeuThrAspIleP 631
 4083 TGACATCTTGGCTGAAGCCCGCAGGAACCTAGACAGATCTTACGAGCAC 4132
 631 roleuThrGlyGlnValaLeuThrTrrPheThrAspLysSerTrrVala 647
 4133 CTCTCCAGACGCTGACACACCTGTGTACCAATGGAGACACCTTCG 4182
 648 ValGlnGlyLysArgMetAlaGlyAlaAlaValaValaAspGlyThrArgTh 664
 4183 CAAGAGGGCGAGCGCAAGGCGGAGCAGCAGTAAACCCAGACCAAGGT 4232
 664 rIleTrrPalaSerSerLeuProGlnGlyTrrSerAlaGlnLysAlaGlnu 681
 4233 AGTGTGGCCAAAGCACCTGCGCAGCGGACATCGGCCCAAGAGCTGAGT 4282
 681 eumeAlaLeuThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleasn 697
 4283 TGAATGCCCTCACCCAGACCTTAAAAATGGCAGAGATGAAGAGTCAAT 4332
 698 IleTrrAspSerArgTrrAlaPheAlaThrAlaHisValHisGlyAl 714
 4333 GTTTCACCGATAGCGCTTATGCTTTGGCCACTCCCATTCACGGAGA 4382
 714 aLleTrrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLysA 731
 4383 AATATATAGAAAGCGCGGGTGTGCTCACATCAAGAAAGAAATCAAAA 4432

```

731 snlysgluGluIleuSerLeuGluAlaLeuHisLeuProLysArg 747
|||||
4433 ATAGAGCAGGATCTTGCCCTACAGGCTCTCTCTGCCCCAAAAGA 4482
748 LeuAlaIleHisCysProGlyHisGlyAlaLysAspLeuIleSe 764
|||||
4483 CTAGAGTAATTCATTCGCCGGACATCGAAGGAAACCGCGCGAGAGC 4532
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaIleLay 781
|||||
4533 AAGGGCAACAGATGGCCGACCAAGCGCCGAGAGTACCACTAGA. 4581
781 alaSnLeuLeuProIleIleGluThrProLys..... 791
|||||
792 .....AlaProGluProArgArg.....GlnThrThrLeuGluAs 803
|||||
4612 GAAATTCAGCCCTATACATGACATTCATATACGAGTACGACTGA 4661
803 pThrGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGly 820
|||||
4662 CATAAAGATCTGACTAACTAGGCGCCACTTATAGCAGATGCAAG...A 4708
820 hrCysTyrThrSerTyrGlyLysGluIleuProHisLysGluGlyLeu 836
|||||
4709 AGTGTGGGTTATCAGGAAAGCCGTAAATGCCGTGATCAATTCACCTTT 4758
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLe 853
|||||
4759 GACATATTAGATTCTTCATCAATTCACCCACTCAGTTCTCAAAAAAC 4808
853 uGlnGlnLeuValArgThrSer.....ProTyrHisValLeuLeuLeu 868
|||||
4809 AAAGCTCTCTCTGAAAGAACTAGCTCTCTTATTCATGCTGAAACGGG 4858
868 roGlyValAlaAspSerValValLysHisCysValProCysGlnLeuVal 884
|||||
4859 ATCGAAGCGTCMAAGACATCAGACTGAGCCCAAGCGTGTGACAGGTC 4908
885 AsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLysSerH 901
|||||
4909 AATGCCAGCAAGTCTGCGCTCAACAAGGACTAGAGTTGAGGCGACCG 4958
901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
|||||
4959 ACCGGGACCCACTGGGAAATTCATTCAGTCTGAGTAAACCTGGCCTGT 5008
918 yTrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTyrVal 934
|||||
5009 ATGGGTATTAATATCTTTAGTTTCATAGACACTTCTCTGGAGGGTA 5058
935 GluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysI 951
|||||
5059 GAAGCTTCCCAACCAAGAAACAACTGCAAGTGTGACCAAGAAACCT 5108
951 eleuGluGluIlePheProArgPheGlyLysLeuProLysValIleGlySerA 968
|||||
5109 ACTAGAGAAATCTTCCCGATTCGGCATGCGACAGGTATTTGGGAACCG 5158
968 sPasnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIle 984
|||||
5159 ACAATGGGCTCTCTCTCTCCAGGTAAGTACAGACGTAGCGCATTTA 5208
985 LeuGlyIleAspTyrLysLeuHisCysAlaTyrArgProGlnSerSerG 1001
|||||
5209 CTGGGGGTGATTTGGAACACTAGTGTCTTACAGACCCAGAGTTTACG 5258
1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
|||||
5259 TCAGGTACAGAAATGAATAGACAAATCAAGAGACTTAACTAAATTTGA 5308
1018 hrThrGluThrGlyIleAsnAspTyrMetAlaLeuLeuProPheValLeu 1034

```

```

|||
5309 CGCTTGCACACTGGCTCTAGGACTGGGTGCTCTCTCTCCCTACCCCTG 5338
1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlyLe 1051
|||||
5359 TATCGAGCCCGACACAGCGCGGGCCCATGAGTGTTCACCCCATATGAAT 5408
1051 uLeuTyrGlyGlyProProProLeuAlaGlnIleAlaPheAlaHisSerA 1068
|||||
5409 CTTATATGGGGACCCCGCCCTTGTAACTCTTCGATCCGATCCGATCGA 5458
1068 lAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
|||||
5459 CAAAGTTACATCATACCCCTCTCTCCAGGCCATTTACAGGCACTTAC 5508
1085 TrpValArgGlnArgAlaTrpLysGlnLeuArgGlnAlaTyrSerGly 1101
|||||
5509 CTGGTCCAGCAGAAAGTCTGGAGACCGTTGGCGGCACTTACCAA...GA 5555
1101 yAspLeuGln.....ValProHisArgPheGlnValGlyAspSerV 1115
|||||
5556 ACAACTGACCGCGCGGAGTGTCTTACCTTCTCCGAGTGGAGACAG 5605
1115 aLTrValAlaArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGly 1131
|||||
5606 TGTGGTCCCGCAGACACCAAACTAAATCTAGAACCCCGCTGGAAGA 5655
1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGly 1148
|||||
5656 CCTTATACCGCTCTTACGACTACCCCGCTCTCAAGTGTGAGCGCAT 5705
1148 eProThrTrpIleHisAlaSerHisValLysProAla..... 1160
|||||
5706 TGCAGCTGTGATCCAGCTCCCGCAAGGCTGCGGACACAGAGATTG 5755
1161 ..ProProProAspSerGlyTyrArgAlaGluLysThrGluAsnProLeu 1176
|||||
5756 AGCCACCATGGAGATGACATGCGCTGTTCAACGCTCAAAATGCCCTA 5805
1177 LysLeuArgLeuHisArg 1182
|||||
5806 AAGATTAAGATTGACCCCG 5823

seq_name: /cgn2-6/ptodata/2/ina/5A_COMB.seq.us-08-110-300A-9
seq_documentation_block:
; Sequence 9, Application US/08110300A
;
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; STREET: 45 Hoxie Faithfull and Hapgood
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,300A
; FILING DATE: 20-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50

```

```
TELECOMMUNICATION INFORMATION
TELEPHONE: 212-757-2200
TELEFAX: 212-566-1461
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10367 base pairs
Type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYDROTHERMAL: NO
ANTI-SENSE: NO
US-08-110-300A-9
```

[illegible]

```
alignment_block:
US-09-171-553B-5 x US-08-110-300A-9/rev .
```

Align seg 1/1 to reverse of: US-08-110-300A-9 from: 1 to: 10367

[illegible]

199	LeuIIeGInGInGlyIIeLeuValProValGInSerProThrAsnThr	215
4760	CTGCTGGATTCAGGGCAATCTGGTCCCTGCACGTCCTCCGGAAACGCC	4711
215	oleuLeuProValArgIysProGlyThrAsnAspTryArgProValGInA	232
4710	CCGCTACCCGTTAAAGAACCGGGGACTAATGATATATAGCCCTGTCCAG	4661
232	spleuAArgIuValAsnIysArgValGInAspIleHisProThrValPro	248
4660	ANCTGAGAAATCAACAAAGCGGGGAGCAACATCCACCACCGTCGCC	4611
249	AsnProTryAsnIleuLeuSerAlaIleuProProGluArgAsnTryIyrTh	265
4610	AACCTTCAACACCTTTCAGCGGGCTCCACCGTCGCCACCAAGTGTACAC	4561
265	rValIeuAspIeuLysAspAlaIlePheCysLeuArgIleuHisProThs	282
4560	TGTCGTTGACTTAAAGATGCTTTTTCGCTGAGACTCCACCCACCA	4511
282	erGInProIeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThr	298
4510	GTCAGCTCTCTTCGCTTGTGATGAGACATCCAGATGTGGAAATGTCA	4461
299	GlyGInLeuThrTrpThrArgIeuProGInGlyPheIysAsnSerProTh	315
4460	GGACATTTAACCTGGACGAGCTCCCGAGGGTTCAAAAACATCCAC	4411
315	rIlePheAspGluAlaLeuHisArgAspIleuAlaAsnPheArgIleGIn	332
4410	CGCTGTGATGAAGCCCTCACAGGACTCCGCAACTTCGGATCCAGC	4361
332	IsProGInValThrIleuGInTryValAspAspIleuLeuAlaGly	348
4360	ACCCAGACCTGATTCGTCCAGATATGATGATGACTTACTCTGGCGCC	4311
349	AlaThrLysGInAspCysLeuGluGlyThrIyrAlaIleuLeuGInIle	365
4310	ACTTCGAGCTACTCTCAACAGGATACGCGGCTCGTTACAAACCT	4261
365	userAspIeuGlyTyrArgAlaSerAlaLysIysIaGInIleCysArgA	382
4260	AGGGGACCTGGATATCGGGCTCGGCCCAAGAACCCAAATTTCCAGA	4211
382	rgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGInArgIreue	398
4210	AACGCTCAGATATCTGGGTATCTTCTAAAGAGGCTACGAGATGGCTG	4161
399	ThrGluAlaArgLysTrpThrValGInIleProAlaProThrThAl	415
4160	ACTAGAGCGCAAAAGACCTGTGATGGGGCACCTACTCGCAACACCC	4111
415	AluGlyValArgIlePheLeuGlyThrAlaGlyPheCysArgIeuTrpI	432
4110	TGCACAATAGGAGATCTTCAGGGAGCGCAGGCTCTGTCCCTGTGA	4061
432	IleProGlyPheAlaThrLeuAlaAlaProIeuIyrProIeuThrIyrGlu	448
4060	TCCCTGGGTTTGCAAAATGGCAGCCCTTGTAACCTCTCCACAAAAG	4011
449	LysGlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaI	465
4010	GGGACTCTGTTGAGTGGGGCCAGACACCAAAAGGCCATCACAGAT	3961
465	eLysLysAlaIeuLeuSerAlaProAlaIeuAlaIeuProAspValThrL	482
3960	CACAGAGGCTCTTAACTGCCCTGCCTGGATGTGCCAGACTTATACA	3911
482	ysProPheThrIleuIyrValAspGluArgGlyValAlaArgGlyVal	498
3910	AGCCCTTGCACTTTTGTATACCAAGACGAGGCTACGCCCAAGGTGTC	3861
499	LeuThrGInThrIleuGlyProTryArgArgProValAlaIyrIleuSerIy	515

```

|||||
3860 CTATGCCAAACATGGGGCTTGGCGTCGGCGGCTGCTACCTCCAA 3811
515 slylsleuaspProvalalaSerGlyTrpProvalaCysleuysAlaIlea 532
3810 AAAGGTAGACCCAGTGGCAGCTGGGTGGCCCTTGCTTACGAGATGGTAG 3761
532 laAlaValaAlaIleuValaIysAspAlaAspLysleuThleuGlyGln 548
3760 CAGCCATCGCGCTCTGACCAAGCGCTGGCAAGCTCACCATGGAGACAG 3711
549 AsnIleThValIleAlaProHisAlaLeuGlnAsnIleValaArgGlnPr 565
3710 CCATAGATCATTTGGCCCCCATGGCATAGACAGCAGTGTAAAGACACC 3661
555 oProspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuL 582
3660 CCGTATGCTGGCTCTCCAAAGCGCCGATGACCCACTACAGAGCTCTGC 3611
582 euLeu...ThrgIaArgValThrpheAlaProAlaAlaLeuAsnPro 597
3610 TTTCGACACGACGACGAGTCCAGTTCGGACCAATAGGGCCCTAAACCCA 3561
598 AlaThrIleuLeuProGluGluThrAspGluProValThrHisAspCysH 614
3560 GCTACGCTGCTCCCT...CTACCTGAGAGAGGGGCTGCACATGACTGCT 3514
614 sGlnLeuLeuIleGluGluThrGlyAlaArgLysAspLeuThrAspIleP 631
3513 TGACATCTTGGCTAGAGCCGACGAGCTAGACAGATCTTACGAGCAGAC 3464
631 roLeuThrGlyGluValLeuThrTrpPheThrAspLysSerSerIyVal 647
3463 CTCGCCAGAGCGCTGACACACCTGCTACACAGATGGAGAGCTTCCG 3414
648 ValGluGlyLysArgMetAlaGlyAlaAlaValaIAspGlyThrArgTh 664
3413 CAAGAGGGGAGCGCAAGCGCGAGAGCATGACACCGAGACCGAGAGGT 3364
664 rIleTrpAlaSerSerIeuProGluGlyThrSerAlaGlnLysAlaGlu 681
3363 AGTCTGGCGCAAGACACTGCCAGCGGAGCATGGGCCCAAGAGCTGAGT 3314
681 euMetAlaLeuThrGlnAlaLeuArgLysAlaGluGlyLysSerIleAsn 697
3313 TGATAGCGCTACCCCAAGCCTTAAATAATGGCAGAGTAAGACCTGAT 3264
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
3263 GTTACACCGATAGCCGTTATGCTTTCCTCCACTGCGCCATATTCACGGAGA 3214
714 AlIeTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysA 731
3213 AATATAATAGAAAGCGCGGGTGTCTACATCAGAAAGGAATAATCAATCAAA 3164
731 snLysGlnGluIleLeuSerIeuGlnAlaLeuHisLeuProLysArg 747
3163 ATAAAGAGAGATCTTGGCCTACTGAAAGGCTCTTCTGCGCCCAAAAGA 3114
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspIleLeu 764
3113 CTTACATATTCATTTGCCCGGAGCATCAGAAAGGAAACCGCGCGAGGC 3064
764 rArgGlyLysGlnMetAlaAspArgValaIaLysGlnAlaIaGlnAlaVal 781
3063 AAGGGGCAACAGATGGCGCCAGCGCGCCGAGAAAGTACCACTAGA. 3015
781 alaAsnLeuLeuProIleIleGluThrProLys..... 791
3014 .....GAAAGTCCAGAGACTTCCACACTTGTGATA 2985
792 .....AlaProGluProArgArg.....GlnTyrThrLeuGlnLys 803
|||||
2984 GAAATATCAGCCCTATATCATGACATTTTCAATATACGTGACGA 2935
803 pTrpGlnGluIleLysIleAspGlnPheSerGluThrProGluGly 820
2934 CATAAAGATCTGACTAAACTAGGGGCCACTTATAGACATGCAAAAG...A 2888
820 hrcCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeu 836
2887 AGTGTGGGTTATTCAGAAAGCGCTGAATGCTGTATCAATTCACCTTT 2838
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrHisLys 853
2837 GAACATATAGATTTTCTATCATTCATACCCACCTCAGTTTTCAAAAAC 2788
853 uGlnIleuValaArgThrSer.....ProTyrHisValLeuArgLeuP 868
2787 AAAGCTCTCTCTAGAAAGAACTACTGTCTCTTATACATGCTTCAACCGG 2738
868 roGlyValaIaAspSerValIValLysHisCysValProCysGlnLeuVal 884
2737 ATGCACAGCTCAAAAGACATCATGAGACTTGGCCAAAGCTGTGCACAGCTC 2688
885 AsnAlaAsnProSerArgIleProGlyLysArgLeuArgLysSerH 901
2687 AATGCCAGCAAGTCTGCCCTCAAAACAGGAGACTAGAGTTCGAGGGCACC 2638
901 sProGlyAlaHisThrGluValaAspPheThrGluValaLysProAlaLysT 918
2637 ACCGGGACCCACTGGGAAATTGATTTTCACTGAGTAAACCTGCGCTGT 2588
918 IyrGlyAsnLysTyrLeuLeuValaPheValaAspThrPheSerGlyTyrAl 934
2587 ATGGGATTAATATCTTTAGTTTTCATAGACACTTCTCTGATGGGTA 2538
935 GluAlaTyrProThrLysLysGluThrSerThrValaIaLysLysI 951
2537 GAAGCTTTCCTCAACCAAGAAAGAACTGCCAAAGCTGTGAACCAAGAACT 2488
951 eleuGlnGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
2487 ACTAGAAAGAAATCTTCCCGAGATTCGSCATGCGCACAGGATATGGGAACG 2438
968 spAsnGlyProAlaPheValaIaGlnValaSerGlnGlyLeuAlaLysIle 984
2437 ACAATGGGCTGCTCTGCTCCCAAGTAAGTACAGACAGTACCGCATTTA 2388
985 leuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerG 1001
2387 TTGGGGGTGATTTGAAACTATGATTTGCTTACAGACCCAGAGTTCAGC 2338
1001 yGlnValaGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuT 1018
2337 TCAGTAGAAAGAAATGATATAGACATCAAGACACTTAACTAAATTTGA 2288
1018 hTrpGluThrGlyLysAsnAspTrpMetAlaLeuLeuProPheValLeu 1034
2287 CGCTTGAAGTGGCTTGAAGGATGGGTGCTGCTGCTTCCCTTGCCTG 2238
1035 pheArgValaArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLe 1051
2237 TATGAGCCGCCACACAGCCGGGCCCCCAGCGTCTCACCCCAATGAAAT 2188
1051 uLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerA 1068
2187 CTATATATGGGACACCCCGCCCTTGAATCTTCCTGATCCGATGATG 2138
1068 laAspValIleLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
2137 CAAGGTACTCATAAACCTCTCCAAAGCTCATATTACAGCAGACTTAC 2088
1085 TrpValaGlnArgAlaTyrLysGlnLeuArgGluAlaIaTyrSerGly 1101
2087 CTGGTCCAGCAGCAAGTCTGGAAGCCGTTGGCGGACGCTTACCAA...CA 2041

```

[illegible]


```

265 rValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrS 282
|||||
4560 TGTGCTTGAAGTAAAGATGCTTTTTCCTGAGACTCCACCCACCA 4511
282 eGlnProPheAlaPheGluThrParAspProGlyThrGlyArgThr 298
|||||
4510 GTGAGTCTCTCTTCGCTTGTGAGTGGAGATCCAGAGATGGGAATCTCA 4461
299 GlyGlnLeuThrThrPheArgLeuProGlnGlyPheLysAsnSerProTh 315
|||||
4460 GGACATTAACTCGACACGACTCCCGAGGCTTCAAAAACAGTCCAC 4411
315 rIlePheAspGluAlaLeuHisArgAspLeuLysAsnPheArgIleGln 332
|||||
4410 CCGTGTGATGAAGCCCTGCACAGGAGCTCCAGACTTCGATCCGATCCAGC 4361
332 AspProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGly 348
|||||
4360 ACCAGACTGATTCCTGCTCCAGTATGTAGATGACTTACTGCTGCCGCC 4311
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuLys 365
|||||
4310 ACTTGTGAGCTTGAAGTCAACAGATGAGCGGCGCTGTTCACAAACCT 4261
365 userAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArg 382
|||||
4260 AAGGAGCTCGGATATCGGCGCTCGCCCAAGAACCCCAATTTCCAGA 4211
382 rGlnValThrTyrLeuGlnGlyTyrSerLeuArgGlyGlyGlnArgThrLeu 398
|||||
4210 AACAGGTCAAGTATCTGGGATCTTTCTAAAGAGAGGTCAAGATGCGCTG 4161
399 ThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAl 415
|||||
4160 ACTGAGGCCAGAAAGAGACTGTGATGGGCGACCTACTCCGAGAACCC 4111
415 aLysGlnValArgLysPheLeuGlnTyrThrAlaGlyPheCysArgLeuPhe 432
|||||
4110 TCGAACACTAAGGAGCTTCTTACGAGGAGGAGGCTTGTGCGCTCTGGA 4061
432 LeuProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlu 448
|||||
4060 TCCCTGGGTTTGCAGAAATGGCAGCCCTGTACTCCCTCACCACCAACG 4011
449 LysGlyGlyPheSerThrAlaProGlnHisGlnLysAlaPheAspAla 465
|||||
4010 GGGACTGTGTTGAGTGGGCGCCAGACCAAGGCGCTTACCAAGAT 3961
465 eLysLysAlaLeuSerAlaProAlaLeuAlaLeuProAspValThrL 482
|||||
3960 CAAAGAGGCTCTTAACTGCCCCCTGGGATTCAGAGACTTGACTGA 3911
482 yspProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyVal 498
|||||
3910 AGCCCTTGAAGCTTTTGTGTGACGAGAGCAGGCTACGCCAAGGTGTC 3861
499 LeuThrGlnThrLeuGlyProThrArgArgProValAlaTyrLeuSerLys 515
|||||
3860 CTTAAGCGCAAACTGGGCGCTTGGGCGCGCGGCTTACCTGCTCA 3811
515 sLysLeuAspProValAlaSerGlyThrProValCysLeuLysAlaLeu 532
|||||
3810 AAAGCTAAACCAAGTGGCAGCTGGGCTGGCCCTTGCCTACGAGTGGAG 3761
532 AlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGln 548
|||||
3760 CAGCATGCGCGTCTGACAAAGCGCTGGCAAGCTCACCATGGGACAG 3711
549 AsnIleThrValIleAlaProHisAlaLeuGlnLysAsnIleValArgGlnPr 565
|||||
3710 CCACAGTCAATTCGCGCCCATGACAGTACAGCAGTAAAGTGAACAC 3661
565 oProAspArgThrPheMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
|||||
3660 CCGTATGCTGAGCTCTCCAAAGCCGCAATGACCCACTACAGAGCTCTGC 3611
582 euleu...ThrGluArgValThrPheAlaProAlaAlaLeuAsnPro 597
|||||
3610 TTCTGGACACGACGACGAGTCCAGTTCGACCAATAGTGGCTTAACCCA 3561
598 AlaThrLeuLeuProGlnGluThrAspGluProValThrHisAspCysH 614
|||||
3560 GCTACGCTGCTCCCT...CTACCTGAGAGAGGGCTGCACATGACTGCTCCT 3514
614 sGlnLeuLeuIleGluGlnThrGlyValArgLysAspLeuThrAspIleP 631
|||||
3513 TGACATCTTGGCTGAAGCCAGGAACTAGACAGATCTTAAGGACGACG 3464
631 rIleThrGlnGlyValLeuThrThrPheThrAspGlySerSerThrVal 647
|||||
3463 CTCCTCCAGAGCTGACACACCTGGTACACAGATGGAGACGACTTCTCT 3414
648 ValGlnGlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgTh 664
|||||
3413 CAAGAGGCGGACGCGCAGCGGAGCAGAGTAAACCCAGACCGAGAGT 3364
664 rIleThrAlaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGlu 681
|||||
3363 AGTGTGGCCAAAGCACTGCGACGCGGACATGCGCCCAAGAGCTGAGT 3314
681 eumetalLeuThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleAsn 697
|||||
3313 TGATAGCGCTCACCCCAAGCTTTAAATGSCAGAGGTAAAGAGCTGAAT 3264
698 IleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
|||||
3263 GTTTACACCGATAGCCGTATGCTTTGCCACAGCCCATATTCACGAGAG 3214
714 alleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLysA 731
|||||
3213 AATATATGAAGGCGCGGTTGCTCACATCAGAAGGAAGAAATCAAAA 3164
731 snLysGlnGluIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArg 747
|||||
3163 ATAAGACGAGATCTTGCCCTACTGAGGCTCTTCTCGCCCAAAAG 3114
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIle 764
|||||
3113 CTTAGCATATTCATTTGCCCGGAGCATCAAGAAAGCGCGGAGAGC 3064
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAla 781
|||||
3063 AAGGCGCAACAGATGGCGCACCAAGCGCGCCGAGAAATGACCTAGA. 3015
781 alaSnLeuLeuProIleIleGluThrProLys... 791
|||||
3014 .....GAAACTCCAGAGACTTCCACACTTCGATA 2985
792 .....AlaProGluProArgArg.....GlnTyrThrLeuGlnAs 803
|||||
2984 GAAATTCACCCCTATCTACTCATGAACATTTTCACTAATACGGTACTGA 2935
803 pTTPGlnGluIleLysLysIleAspGlnPheSerGlnThrProGlnGlyT 820
|||||
2934 CATTAAGATCTGACTTAACCTAGGGCCACTTATGAGATGCAAAAG...A 2888
820 hrcYsTyrThrSerTyrGlyLysGluIleLeuProHisLysGlnGlyLeu 836
|||||
2887 AGGTGTGGTATTCAGAGGAAAGCTGTATATGCTATCAATTCACCTTT 2838
837 GluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisIle 853
|||||
2837 GAACATTAATGATTTCTCATCAATGACCAACCTCAGTTTGCACAAAC 2788
853 uGlnGlnLeuValArgThrSer.....ProTyrHisValLeuAlaGluLeuP 868
|||||

```

```

2787 AAAGCTCTTCTAGAAAGAACTACTGCTTATTACATGCTGAACCGG 2738
868 roglYValAlaAspSerValValLysHIScysVal)ProCysGlnLeuVal 884
2737 ATCGAAGCTCAAGACATCACTGAGACTGGCCAGGCTGTGCACAGGTC 2688
885 AsnAlaAspProSerArgIleProProGlyLysArgLeuArgGlySerH 901
2687 AATGCCAGAGAGTCTGCGTCAACAGGAGACTAGAGTTGAGGGCACCG 2638
901 sProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysT 918
2637 ACCCGGACCCACTGGGAAATGATTCTCAAGTAGTAAACCTGGCCTT 2588
918 YrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 934
2587 ATGGGATATAAATCTTTAGTTTCATAGACACTTCTCGTGATGGGTA 2538
935 GluAlaTyr-ProThrLysLysGluThrSerThrValAlaLysLysI 951
2537 GAAGCTTCCCAACCAAGAAAGAACTGCCAAAGTTGTAACCAAGAGCT 2488
951 eLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
2487 ACTAGAGAAATCTCCCGAGATTGGCATGCCACAGGATTTGGAAACG 2438
968 sPasnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysI 984
2437 ACAATGGGCTGCTGCTGCTCCAGAGTAGTCAGACAGTAGCGCATTTA 2388
985 LeuGlyLysAspTyrLysLeuHIScysAlaTyrArgProGlnSerG 1001
2387 TTGGGGGCTGATTGGAAACTACATTGCTTACAGACCCCGAGGTTGAG 2338
1001 YGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
2337 TCGAGTAGAAGATGATGACATCAAGGAGACTTTTACATAAATTGA 2288
1018 hrThrGluThrGlyLysAsnAspTyrMetAlaLeuLeuProPheValLeu 1034
2287 CGCTTGCACTGGCTAGGAGACTGGGTGCTCCGCTTCCCTGGCCTG 2238
1035 PheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlu 1051
2237 TATGAGACCGCGCAACAGCGCGGCCCGCCAGGCTCACCCTATATGAAT 2188
1051 uLeuTyrGlyLysProProLeuAlaGluIleAlaPheAlaHisSerA 1068
2187 CTATATATGGGGACACCCCGCTTGTAAACTTCCCTGATCCGACATGG 2138
1068 LaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGlu 1084
2137 CAAGGTTACTCATACCCCTCTCCAGCTCATTTACAGGCACTGTAC 2088
1085 TrpValArgGlnArgAlaTyrGlnLeuArgGluAlaTyrSerGly 1101
2087 CTGGTCCAGCAGAGTCTGGACCGTGGCGGACGCTTACCA...GA 2041
1101 YAspLeuGln.....ValProHisArgPheGlnValGlyAspSerV 1115
2040 GCACATAGACCGCGCGGTAGTACTCAACCTTCCGGGTGGTGACACAG 1991
1115 aTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTyrLysGly 1131
1990 TGTGGTCCGACAGACCACTAAATCTAGAAACCCCGCTGGAAGGA 1941
1132 ProTyrLeuValLeuLeuThrThrProThrAlaValLysValGlyI 1148
1940 CCTTATACCGTCTACTGACTACCCCGCCGCTCTCAAAAGTAGACGAT 1891
1148 eProThrTyrPheHisAlaSerHisValLysProAla..... 1160
1890 TGCACGCTGATCCACGCTGCCACGTAAGGCTGCCGACACACAGATTG 1841

```

```

1161 ..ProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeu 1176
1840 AGCCACCGACGAGAAATCGACATGGCGTGTTCACCGTCCCAAAATCCCTA 1791
1177 LysLeuArgLeuHisArg 1182
1790 AAGATTAACATGACCCCG 1773
seq_name: /cgn2.6/prodata/2/ina/PCTUS_COMB.seq:PCT-US93-08041-9

```

```
seq_documentation_block:

```

```
Sequence 9, Application PC/TUS9308041
GENERAL INFORMATION:

```

```

APPLICANT: The Public Health Research Institute of the City of
APPLICANT: New York, Inc.
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Davis Hoxie Faithful and Hapgood
STREET: 45 Rockefeller Pl.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10111

```

```
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: IBM PC compatible
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08041
FILING DATE:

```

```
CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,100
FILING DATE: 28-AUG-1992

```

```
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/____
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:

```

```

NAME: Jacobs, Seth H.
REGISTRATION NUMBER: 32,140
REFERENCE/DOCKET NUMBER: 11698A50
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461

```

```
INFORMATION FOR SEQ ID NO: 9:

```

```
SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 10367 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```

```

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

```

```
PCT-US93-08041-9
```

```
alignment_scores:

```

```

Quality: 3980.00 Length: 1206
Ratio: 3.992 Gaps: 12
Percent Similarity: 82.670 Percent Identity: 62.687

```

```
alignment_block:

```

```

US-09-171-553b-5 x PCT-US93-08041-9/rev
Align seg 1/1 to reverse of: PCT-US93-08041-9 from: 1 to: 10367

```

```

1 GYArgArgGlySerAspProLeuProGluProArgValThrLeuLysVal 17
111 ?:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5360 GGAGCTGAGGCTCAGAGACCGCCCGCTGAAACCCAGATACCTCAGAGT 5311

```



```

3513 TGACATCTTGCGTAGAGCCACGAGACTAGACGATCTTACGGACGAC 3464
631 rolethrlgylgluValleuthrphethrphaserSerTyrVal 647
3463 CTCTCCAGAGCGCTGACCCACCTGGTACACAGTGGAGACGCTTCTG 3414
648 ValgluglylserMetalaValAlaValAlaValaspgllythrArgth 664
3413 CAAGAGGGGCGAGCGGAGGAGCAGTAAACCCAGACCGAGAGGT 3364
664 rtlethrlaserSerleuProgluglylthrSerAlaValaspgllyl 681
3363 AGTGTGGGCGAAGACACTGCGCGGAGCATCGCCCAAGAGTGTGT 3314
681 eumetaleuthrlglnAlaValaspgllylgluValaspSerlleasn 697
3313 TGATAGCGCTGACCCACGCTTAAATAATGGCAGAGGTAAAGAGTGTAT 3264
698 lleytyrthrAspSerArgTyrAlaPheAlaThrAlaHisValHisgluAl 714
3263 GTTACACCGATAGCGCTTATGCTTTTCCACTGCCATATTCACGAGGA 3214
714 allelytyrlysglnArggluValleuthrSerAlaValaspgllyl 731
3213 AATATATAGAGGCGCGGGGTCTCTCATCAGAAAGAAAGAAATCAAAA 3164
731 snlysglnleuthrleuSerleuValAlaValaspgllyl 747
3163 ATATATATAGAGGATCTGGCCCTACTGAAGGCTCTCTCCGCCCAAGGA 3114
748 leuAlaIleleuHisCysProgluHisgluValaspSerlleSe 764
3113 CTTAGCATATTCATTCGCGGAGCATCAGAGGAGAAACCGCGGAGGC 3064
764 rArggluAsnGlnMetalaAspArgValAlaValaspgllylAlaVal 781
3063 AAGGGCCACAGATGGCGGACCGGCGGAGGAGTAAAGCCACTAGA. 3015
781 alaSnleuProleuIlegluThrProlys..... 791
3014 .....GAAACCTCAGAGACTTCCACACTCTGATA 2985
792 .....AlarProIleProArgArg.....GlnTyrThrleuGlnAs 803
2984 GAAATATCAGCCCCCTATCTCTCATGACATTTCTACTATACGTTGACTGA 2935
803 pTyrGlnGluIleLysLysIleaspGlnPheSerGluThrProgluGlyT 820
2934 CATAAAGATCTGATAAAGTAAAGGCGCACTTATGACGATGCAAG...A 2888
820 hcrystyrthrSerTyrGlyLysGluIleuProHisLysgluGlyLeu 836
2887 AGGTGGGTGTTATCAGGGAACCTGTATGCTGATCATATTCACCTTT 2838
837 GluTyrValGlnGlnIleHisArgLeuThrHisleuGlyThrLysHisle 853
2837 GAATATATATATTTCTTCTCATTCATTCAGCCACTCAGTTTCTCAAAAC 2788
853 uGlnGlnleuValArgThrSer.....ProTyrHisValleuArgleuP 868
2787 AAAGGCTCTCTAGAAAGAAAGAACTGCTCTTATATATCTGAAACCGGG 2738
868 rogluValAlaAspSerValValLysHisCysValProCysGlnleuVal 884
2737 ATCAGACGCTCAAGACATCATGACACTTGCCCAAGCCTGACACAGCT 2688
885 AsnAlaAsnProSerArgIleProProgluLysArgleuArgGlySerH 901
2687 AATGCGACAGCTCTGCCGTCAAAACAGGAGCTAGAGTGCAGGCGACCG 2638
901 sProgluValHisTyrpGluValaspPheThrGluValLysProAlaLysT 918
2637 ACCGGGACCACTGGGAATTCATTCCTACGTAGAGTAAACCTGGCTGT 2588

918 yrcLysAsnLysTyrleuLeuValPheValaspThrPheSerGlyTyrVal 934
2587 ATGGGTATATAATTCCTTTAGTTTCATATGACCTTTCTCTGATGGGTA 2538
935 GluAlaTyrProThrLysLysGluThrSerThrValAlaLysLysI 951
2537 GAAGCTTCCCAACCAAGAAAGAACTGCCAAAGTGTAAACCAAGAGCT 2488
951 eleuGlnGluIlePheProArgPheGlyIleProLysValIleGlySerA 968
2487 ACTAGAAGAAATCTCCCGATTCGCGATGCGACAGTATTTGGAAACCG 2438
968 spasnGlyProAlaPheValAlaGlnValSerGlnleuAlaLysIle 984
2437 ACAATGGGCTGCTTCTGCTCCAGTAGTACAGACAGTACGAGTATTA 2388
985 leuGlyIleAspTyrPylLysleuHisCysAlaTyrArgProGlnSerG 1001
2387 TTGGGGGTGATGGAAACTACATTTGTCTTACAGACCCCAAGATTCAGG 2238
1001 yGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 1018
2337 TCAGGTGAAGAATGATGATAGGACATCAAGAGACTTTAAGTAAATGA 2288
1018 hrThrGluThrGlyIleAsnAspTyrMetalaValleuProPheValLeu 1034
2287 CGCTTGACACAGCGCTTAGGAGACTGGGTGCTGCTTCCCTGGCCCTG 2238
1035 PheArgValArgAsnThrProgluGlnPheGlyLeuThrProTyrGly 1051
2237 TATGAGGCGCCCAACACGCGCGGCCCGCCAGGTCTCCACCATATGAAAT 2188
1051 uLeuTyrGlyLysProProleuAlaGluIleAlaPheAlaHisSerA 1068
2187 CTATATAGGGGACACCCCGCCCTTGAATCTCCGATCCGATCGACATGG 2138
1068 laAspValleuLeuSerGlnProleuPheSerArgleuLysAlaValleu 1084
2137 CAAGGTATCATATACCCCTCTCTCCAAAGCTCATTTACAGGCACTGAC 2088
1085 trpValArgGlnArgAlaIleThrLysGlnleuArgAlaTyrSerGly 1101
2087 CTGTCTCAGCAGAGAGTCTGAGAGACGTTGGCGGAGCTTACCAA...GA 2041
1101 yAspLeuGln.....ValProHisArgPheGlnValAlaLysPseV 1115
2040 GCAACTAGACCGCGCGGTAGTACTCTACCCCTTTCGCGGTGACACAG 1991
1115 alTyrValArgArgHisArgAlaGlyAsnleuGluThrArgTyrLysGly 1131
1990 TGTGGGTGCGCAGACCAAACTAAATCTAGAAACCCCGCTGGAAGA 1941
1132 ProTyrleuValleuLeuThrThrProThrAlaValLysValGluGly 1148
1940 CCTATACGCTGCTACTGACTACCCCAACCGCTGCAAAAGTAGAGCAT 1891
1148 eProThrTyrIleHisAlaSerHisValLysProAla..... 1160
1890 TGCAGCGGTGATCCAGCGTCCGACGTAAAGGCTCCGACACAGAGATTG 1841
1161 ..ProProAspSerGlyTyrArgAlaGluLysThrGlnAsnProleu 1176
1840 AGCCACACACAGATGACATGAGCGGTTCACACGCTCCCAAAATCCCGTA 1791
1177 LysleuArgleuHisArg 1182
1790 AAGATAAGATTGACCCGC 1773

seq_name: /cgn2_6/plodata/2/ina/6b_COMB.seq:us-09-011-745-3
seq_documentation_block:
; Sequence 3, Application US/09011745

```

```

? Patent No.6165715
? GENERAL INFORMATION:
? APPLICANT: Collins, Mary KL
? APPLICANT: Weiss, Robin A
? APPLICANT: Takeuchi, Yasuhiro
? APPLICANT: Cosset, Francois-Lolc
? TITLE OF INVENTION: Expression systems
? FILE REFERENCE: 09/011,745
? CURRENT APPLICATION NUMBER: US/09/011,745
? EARLIER FILING DATE: 1998-06-22
? EARLIER APPLICATION NUMBER: PCT/G996/02061
? EARLIER FILING DATE: 1996-08-23
? EARLIER APPLICATION NUMBER: GB9517263.1
? EARLIER FILING DATE: 1995-08-23
? NUMBER OF SEQ. ID NOS: 29
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 3
? LENGTH: 7308
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Portion of
US-09-011-745-3

```

alignment_scores:	
Quality:	3967.00
Ratio:	3.959
Percent Similarity:	82.810
alignment_block:	
Length:	1210
Gaps:	13
Percent Identity:	62.479

alignment_block:
US-09-171-553B-5 x US-09-011-745-3

Align seg 1/1 to: US-09-011-745-3 from: 1 to: 7308

1 GtYArgrtGlyLeuSerProLeuProGluProAlaValThrLeuLysIsta 17
 3184 GGAGGTCAGGGTCAGAGAGCCCCCTTGAAACCCAGATTAACCTCAAGT 3233
 17 IgluGlyGlnProValGluPheLeuValAspThrGluIstiHisSerV 34
 3234 CGGGGGGCAACCCGTCACCTCTCGTGAATACCTGGGGCCCAACACTCCG 3283
 34 alleuLeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMet 50
 3284 TGTGTGACCCAAAAATCTGTGACCCCTTAAGATGAATACCTGCGGGTCCA 3333
 51 GtYAlaThrGlyGlnArgLntTrpOrnThrThrArgAlaThrValAs 67
 3334 GGGGTACTAGGAGAAAGGGGTATGCTGTGGACACAGATGCCAAAGTTC 3383
 67 pleuGlyValGlyAlaValThrHisSerPheLeuValIleProGluCysP 84
 3384 TCTAGGTACCGGTAAAGTCAACCACTCTTCCTCATGTATACCAACACTGC 3433
 84 rvaProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100
 3434 CCTATCTCTGTGTAGGAAGAGATGTCGACTAAACATAAAGCCCAATC 3483
 101 SerPheGlu...GlnGtYArGProGluIstiSerValAsnAsnLysProI 116
 3484 CACTTTGAGGGATCAGGAGCTCAGGTTATGTGGACCAATGTGGGGCCCT 3533
 116 eThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProG 133
 3534 GCAGATGTGTGACCTTAATATAGAAGATGAGACATCGGCTACATGAGCCT 3583
 133 lnaValLysProAspGlnAspIleGlnSer...TrpLeuGluGlnPhePro 148
 3584 CAAAAGACCCAGATGTTCTCTCAAGGCCACATGCTGCTGTGATTTTCC 3633
 149 GlnAlaTrpAlaGluThrAlaGlyMetCylLeuAlaLysGlnValProP 165

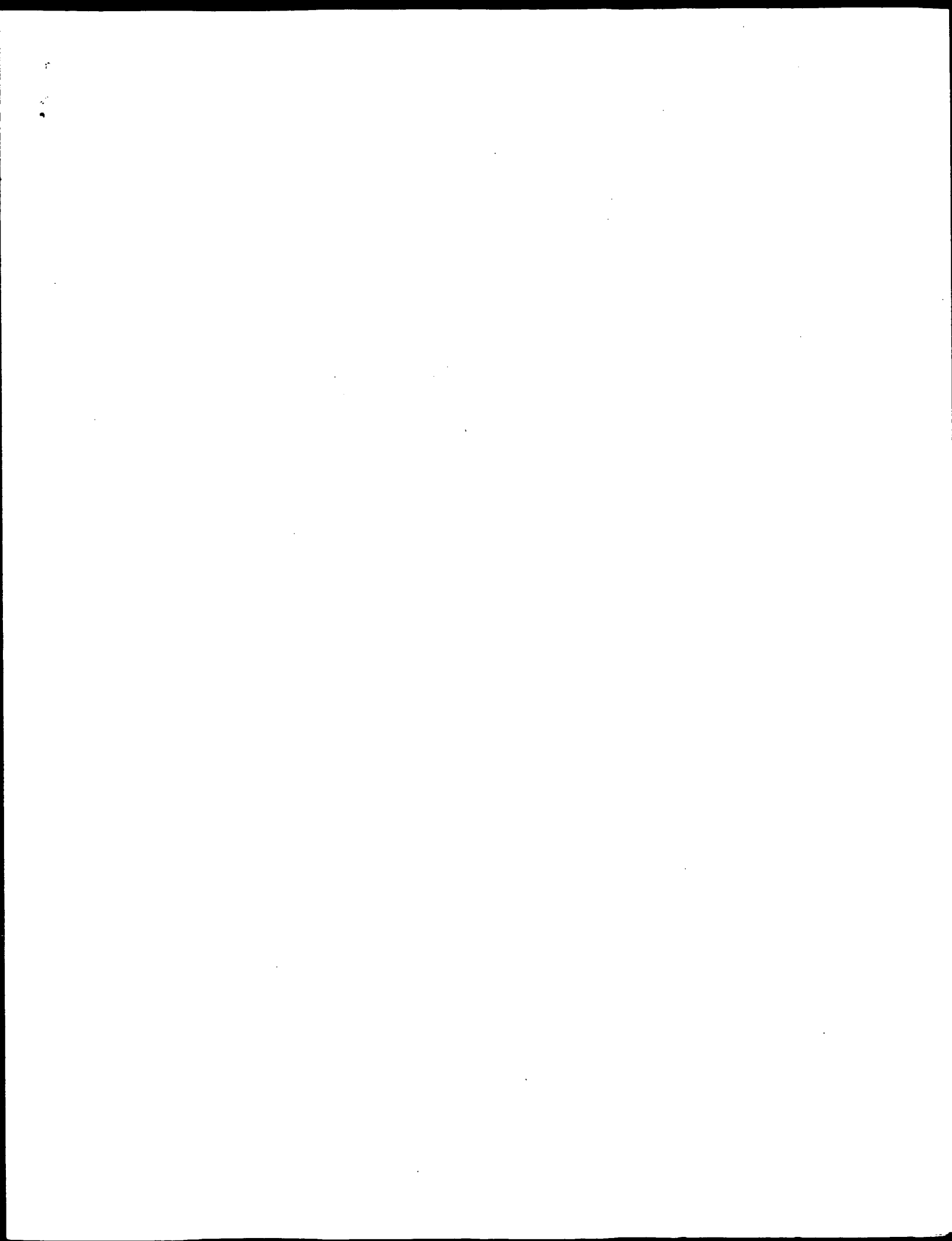
[illegible]

4034 GTCAGCTCTCTCCCTTTGAGTGGAGAGATCCAGAGATGGGAATCTCA 4083
299 GlyLeuThrLeuThrPheArgLeuProGlnGlyPheLysAsnSerProth 315
|||||
4084 GGACAAATTGACCTGGACGAGACTCCACAGGGTTTCAAAAACAGCTCCAC 4133
315 rLePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArgIleGln 332
|||||
4134 CCGTTTGTATGAGCACTGCACAGAGACTGACAGACTTCGGATCCAGC 4183
332 iSProGlnValThrLeuLeuGlnThrValAlaAspLeuLeuAlaGly 348
|||||
4184 ACCGAGACTGATCTCTGCTACAGTACGTGATGACTTCTGCTGCGCCGCC 4233
349 AlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLe 365
|||||
4234 ACTCTGAGCTATGACTGCCAACAAGTACTCGGGCCCTGTTACAAACCT 4283
365 uSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArg 382
|||||
4284 AGGGAACCTCGGGATCGGGCTCGGCCAAGAAAGCCAAATTTGCCAGA 4333
382 rGlnValThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrLeu 398
|||||
4334 AACAGCTCAAGTATCTGGGGTATCTTAAAGAGGTCAGAGATGGCTG 4383
399 ThrGlnAlaArgLysLysThrValAlaGlnIleProAlaProThrAl 415
|||||
4384 ACTGAGGCCAGAAAGAGACTGTGATGGGGCAGCTACTCCGAGAACCCC 4433
415 aLysGlnValArgLysPheLeuGlyThrAlaGlyPheCysArgLeuTyr 432
|||||
4434 TCGACAACTAAGGAGGTTCTTAGGAGCGCAGGCTTGTGCGCTCTGA 4483
432 lProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGln 448
|||||
4484 TCCCTGGTTTCAGAAATGCGACGCCCTTGTACCTCTCCACAAACG 4533
449 LysGlyGlyPheSerThrAlaProGlnHisGlnLysAlaPheAspAla 465
|||||
4534 GGGACTCTGTTAATTGGGGCCAGACCAACAAAGGCCCTATCAAGAAAT 4583
465 eLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThr 482
|||||
4584 CAAGCAAGCTCTTAACTGACCCAGCCCTGGGGTTGCCAGATTGGACTA 4633
482 ySProPheThrLeuThrValAlaAspGlnArgLysGlyValAlaArgGly 498
|||||
4634 AGCCCTTGAATCTTTGTCGACGAGACAGGCGCTACGCCAAAGSTGTC 4683
499 LeuThrGlnThrLeuGlyProTyrArgArgProValAlaTyrLeuSerLys 515
|||||
4684 CTAAACCAAAACTGGGACCTTGGCGTGGCGGCTGCTACCTGCTCAAA 4733
515 sLysLysAspProValAlaSerGlyTyrProValCysLeuLysAlaIle 532
|||||
4734 AAAGCTAGACCCAGTAGCAGCTGGGTGGCCCTTGCCTAGCGATGGTAG 4783
532 lAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGln 548
|||||
4784 CAGCCATTGCGCTACTGACAAAGATGCGAGCAAGCTAATGCGACAG 4833
549 AsnIleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnPr 565
|||||
4834 CCACTAGTCACTCTGCGCCCCCATGCAAGTAGAGGCACTAGTCAAAACACC 4883
565 oProAspArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSerLeu 582
|||||
4884 CCGGACCGCTGGCTTCCAAAGCCCGGATGACTCACTATCAGGCTTGC 4933
582 eLeu...ThrGlnArgValThrPheAlaProProAlaAlaLeuAsnPro 597
|||||
4934 TTTTGGACAGAGACGGGCTCCAGTTGCGACCGGTGGTAGGCCCTGAAACCG 4983
598 AlaThrLeuLeuProGlnGlnThrAspGlnProValThrHisAspCysHis 614
|||||
4984 GCTACGCTGTCTCCA...CTGCTGAGGAAGAGGCTGCACACACATCGCT 5030
614 sGlnLeuLeuIleGlnGlnThrGlyValArgLysAspLeuThrAspIle 631
|||||
5031 TGATATCTGTGGCCGAAGCCCGCAAGCCGACCTTAACCGACACAG 5080
631 rLeuThrGlyGlnValLeuThrTyrPheThrAspLysSerTyrVal 647
|||||
5081 CGCTCCAGACGCCGACACACCTGTGATCAGGATGGAAGACATGCTTGA 5130
648 ValGlnGlyLysArgMetAlaGlyAlaAlaValAlaAspLysThrArg 664
|||||
5131 CAGAGGAGACAGCGTAAAGCGCGAGAGCTCGGTGACCAACCGACCGAG 5180
664 rIleThrAlaSerSerLeuProGlnGlyThrSerAlaGlnLysAlaGln 681
|||||
5181 AATCTGGCTAAAGCCCTGCCAGCGCGGACATCCGTCAGCGGCTGAAAC 5230
681 eMetAlaLeuThrGlnAlaLeuArgLeuAlaGlnGlyLysSerIleAsn 697
|||||
5231 TGATAGCAGTCCAGCCAGCGCTTAAAGATGGCAGAGGTAAGAGCTAAAT 5280
698 lLeuThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAl 714
|||||
5281 GTTATATCTGATAGCCGTTATGCTTGTGCTAGTCCCATTCATGATGAGA 5330
714 aIleTyrLysGlnArgLysLeuLeuThrSerAlaGlyArgGlnLysAl 731
|||||
5331 AATATACAGAAAGCGCTGGTGTGCATCATCAGAAAGAGATCAAAA 5380
731 sLysGlnGlnIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArg 747
|||||
5381 ATAAAGCAGAGATCTGGCCCTACTAAAGCCCTTGTGCGCCAAAGA 5430
748 LeuAlaIleIleHisCysProGlnHisGlnLysAlaLysAspLeuIle 764
|||||
5431 CTTAGCATATTCATTTGTCGAGACATCAAAAGGACACAGCGCGAGGC 5480
764 rArgLysAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAla 781
|||||
5481 TAGAGGCAACCGGATGGCTGACCAAGCGGCCGCAAGGACAGC... 5523
781 aLysLeuLeuProIleIleGlnThrProLysAla... 792
5524ATCAGAGAGACTCCAGACACTCTACCTCTCATATA 5559
793ProGlnProArgArgGlnTyrThrLeuGlnAs 803
5560 GAAATTTGATACCTTACCTCAGACCTCAGAACATTTTCATTAACAGAGACGA 5609
803 rTyrGlnGlnIleLysLysIleAspGln...PheSerGlnThrProGln 819
|||||
5610 TATAAGAGCAACCAAGATTGGGGCCATTATGATTAACAAAGAAAGT 5659
819 lYThrCysTyrThrSerTyrGlnLysGlnIleLeuProHisLysGlnGly 835
|||||
5660 ATTGGGCTTAC...CAAGAAACCTGTGATGCTGACGACAGTTTACT 5703
836 LeuGlnTyrValGlnGlnIleHisArgLeuThrHisLeuGlnTyrHis 852
|||||
5704 TTTGAAATTATAGACTTTCTTCATCAGCTGATGATCAGCTCTCTCAAA 5753
852 sLeuGlnGlnLeuValArgThr...SerProTyrHisValLeuArgL 867
|||||
5754 AATGAAGCTCTCCTAGAGAGAGACCCAGACCTTACATACATGCTGAACC 5803
867 eUProGlnValAlaAspSerValValLysHisCysValProCysGlnLeu 883
|||||
5804 GGGATCGAACACTCAAAATATCAGTGAAGACTGCAAAAGCTTGTGCACAA 5853


```

731 snlysgluuileuSerLeuGluAlaLeuHisLeuProIysArg 747
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4542 ATAAAGACGAGATCTGGCCCTACTAAAGCCCTCTTCTGCCAATAA 4591
748 LeuAlaIleIleHisCysProGlyHisGlnLysAlaLysAspLeuIle 764
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4592 CTTAGCATATTCATTGTCAGACATCAAAAGGACACAGCGCGAGGC 4641
764 rArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaVal 781
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4642 TAGAGGCAACCGGATGCTGACCAAGCGCCGAAAGGCAAGC..... 4684
781 alaSnLeuLeuProIleIleGluThrProLysAla..... 792
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4685 .....ATCACAGACACTCCAGACACCTCTACCTCTCTCATATA 4720
793 .....ProGluProArgArgGlnThrThrLeuGlnLys 803
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4721 GAATAATTCATCACCTTACACCTTCAGACATTTTCATTACACAGTCTGA 4770
803 pTrpGlnGluIleLysLysIleAspGln...PheSerGluThrProGluG 819
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4771 TATAAAGACCTAACCAAGTTGGGGCCATTTATGATAAACAAGAAGT 4820
819 LThrCysThrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 835
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4821 ATTGGGCTTAC.....CAAGAAACCTGTGATGCTGACCAAGTTTACT 4864
836 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHis 852
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4865 TTGGAATTTTATGACTTTTCTTCACTGACTGACTCAGCTGAGCTTCATA 4914
852 sLeuGlnGlnLeuValArgThr.....SerProTyrHisValLeuArgL 867
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4915 AATGAAGGCTCTCCCTAGAGAGAACCCAGCTCCCTACTACATGATGAAACC 4964
867 euProGlyValAlaAspSerValValLysHisCysValProCysGlnLeu 883
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4965 GGGATGCAACACTCAAAATATCCTGACCTGCAAGCTGTGTGCACAA 5014
884 ValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgLys 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5015 GTCAAGCCGACAGCAAGTCTCCCTTAAACAGGAACTAGGGTCCGGGGCA 5064
900 rHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaL 917
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5065 TCGGCGCCGGCAGCTCATTTGGAGATGATTTCAACGAGATAAGCCCGGAT 5114
917 ystTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrp 933
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5115 TGTATGGCTATATAATATCTTCTAGTTTATATGATATACCTTTCTGGCTGG 5164
934 ValGluAlaTyrProThrLysLysGluThrSerThrValAlaLysLys 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5165 ATAGAAGCCTTCCCAACCAAGAAAGAAACCCGAAAGTGTAACCAAGAA 5214
950 sIleLeuGlnGluIlePheProArgPheGlyThrProLysValIleGly 967
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5215 GCTACTAGAGAGATCTTCCAGGTGCGCATGCTCAGATTTGGGAA 5264
967 eAspAsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLys 983
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5265 CTGACAAATGGCGCTTCTGCTCCAAAGGTAGTACAGACAGTGGCGGAT 5314
984 IleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSer 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5315 CTGTGGGGGATTTGGAATTTACATTGTGATACAGACCCCAAGCTC 5364
1000 rGlyLysValGluArgMetAsnArgThrIleLysGluThrLeuThrLysL 1017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5365 AGGCCAGGTAGAAAGATGAATGAAACCATCAGAGAGACTTTAACTAAT 5414
1017 euThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheVal 1033
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5415 TAACGCTTGCAACTGGCTCTAGAGACTGGGTGCTCTACTCTCCCTTAGCC 5464
1034 LeuPheArgValArgAsnThrThrProGlyGlnPheGlyLeuThrProTyrGly 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5465 CTGTACCGAGCCCGACACCGCCCGGCCCCCATGGCTCACCCCATATAGA 5514
1050 uLeuLeuTyrGlyGlyProProProLeuValGluIleAlaPheAlaHis 1067
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5515 GATCTTATATGGGGAGACCCCGCCCTGTATAACTTCCCTGACCTGACA 5564
1067 eAlaAspValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeu 1083
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5565 TGACAAAGATTACTTACACGCCCTCTCTCCAAAGCTCACTTACAGGCTCTC 5614
1084 GluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSerGly 1100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5615 TACTTAGTCCAGCAGAAAGTGGAGACCTCTGGCGGCGGACTTACCAA.. 5662
1100 yGlyAspLeuGln.....ValProHisArgPheGlnValGlyAsp 1114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5663 .GAACAACTGGACCGACCGGAGTGTACTTCAACCTTACCGAGTGGCGACA 5711
1114 eValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLys 1130
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5712 CAGTGTGGTCCGCCGACACAGACTAAGAACTAGAAACCTCGCTGGAAA 5761
1131 GlyProTyrLeuValLeuLeuThrThrThrProThrAlaValLysValGlu 1147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5762 GGACCTTACACAGTCTGCTGCTGACACACCCCAACCCCTCAAAAGTAGACGG 5811
1147 yIleProThrTrpIleHisAlaSerHisValLysProAlaProPro... 1162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5812 CATCGACGCTTGATATACAGCGCGCCACGTGAAGGCTGCCGACCCGGGG 5861
1163 .....ProAspSer.....GlyTrpArgAlaGluLysThrGluAsnPro 1175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5862 GTGGACCATCTCTTGAAGTGCATGATGGCGGCTTCAACGCTCTCAAAACCCC 5911
1176 LeuLysLeuArgGlnHisArgLeuValPro 1185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5912 TTAATAATTAAGTTTAACCCCGGAGGCCCCC 5941

```


```

1311 GGAGCTGCGTCACCTCCACGATGAGACTGGAAATGGCCGATCTCTTC 1360
151 GlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTyr 167
1361 CAGGACCGGGTAAATTTCTCTTGTTCATTCGCGCGCGGAGACAGACAA 1410
167 SmetMetLysLeuTyrLysAspLysSerProSerAspLeuAspT 184
1411 AGTAGAGAACTATATAAAGATAGAGCTGCCATCCATCAGACTTAGATT 1460
184 TyrLeuLysIleSerPheThrGlu..ArgLysThrGlyLysTyrSerLys 200
1461 ATCTAAAGATTAAGTTTCACTGAAAAAGAAACAGGAAATATTCAAAAG 1510
200 AlaLysTyrPyrGluLeuGlyAsnSerPheLeu..LeuTyrGlyGly 216
1511 TGGATTAATAGTATGAGCTGGGGAATAGTTTTTATTAATATGCGGGGG 1560
216 ValAspLysSerThrLeuThrIleArgLeuArgIleGluThrGlyThr 233
1561 AGCAGGGTCCACTTAACCATTCGCTTAGATAGAGAGGAGGACAGAAC 1610
233 ProProValAlaMetGlyProAspLysValLeuAlaGluGluGlyPro 249
1611 CCCCTGTGCGACGTGGAGCCCGATTAAGTACTGGCTGAACAGGGGCCCG 1660
250 AlaLeuGluProProHisAsnLeuProValProGluLeuThrSerLeu 266
1661 GCCCTGGAGCCACCGCATTAACCTGCGGTCGCCCAATTAACCTCGCTCG 1710
266 GProAspLysThrGlnProProSerAsnSerThrGlyLeuIleProT 283
1711 GCTTACATTAACACAGCCGCTTAGCACAGCTGATGATTGATTCCTA 1760
283 HisAsnThrProArgAsnSerProGlyValProValLysThrGlyGln 299
1761 CCACACGCCCTTGAAGACTCCCGAGTGTTCCTGTTAAGACAGGACAGGA 1810
300 LeuPheSerLeuIleGluGlyAlaPheGlnAlaIleAsnSerThrAsp 316
1811 CTCTTACAGTCTCATCCAGGAGCTTCCACACCATCACTCCACGACGCC 1860
316 CAspAlaHisSerSerCysTyrPheCysLeuSerSerGlyProProTyr 333
1861 TGATGCCACTTCTCTTGTGCTTGTGCTTATCCTCAGGCGCTCTTATT 1910
333 YrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHis 349
1911 ATGAGGAGATGCTTAAGAAAGAAATTCATGTGACCAAGACATAGA 1960
350 AsnGlnCysThrTyrGlySerArgAsnLysLeuThrLeuThrGluVal 366
1961 AATCAATGTACATGAGGGGTCCCGAAATTAAGCTTACCTCACTGAAGTTC 2010
366 rGlyLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisLeu 383
2011 CGGGAAGGAGCATGATAGGAAAGCTCCCATCCACCAACACCTT 2060
383 ySTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnTyrLeuVal 399
2061 GCTATAGACTGTGTTATGAGAGGCTCAGAAATTCAGTATTATTAAGTA 2110
400 ProGlyTyrAsnArgTyrPheAlaCysAsnThrGlyLeuThrProCys 416
2111 CCTGTTATAACAGGTGGGTCATGCAATACAGGGGTTAACCCCTGTGT 2160
416 lSerThrSerValPheAsnGlnSerLysAspPheCysValMetValGln 433
2161 TTCACCTCAGTCTTCAACCAATCCAAAGATTTCTGTGATGATGTCAAA 2210
433 lValAlaProArgValTyrTyrHisProGluGluValAlaLeuAspGlu 449
|||||
2211 TCGTCCCCCGAGTGTACTACCATCTGAGAGAGTGTCTTGTATGAATAT 2260
450 AspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAl 466
2261 GACTATCGGTATTAACCGCAAAAAGAAAGAACCCGTATCCCTTACCCTAG 2310
466 aValMetLeuGlyLeuGlyThrAlaValAlaGlyValGlyThrGlyThr 483
2311 TGTATGCTGGATTTAGGAGAGCGCCGTTGGCTGTGAAACAGGACACTG 2360
483 lAlaLeuIleThrGlyProGlnGlnLeuGlyLysGlyLeuGlyLeuHis 499
2361 CCTGATTCACAGGACACAGCAGCTAGAGAAAGCACTTGGTAGCTACAT 2410
500 AlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsn 516
2411 GCGGCCATGACAGAGATCTCCGAGCCTTAGAGAGTCTGTGACCAACT 2460
516 uGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArg 533
2461 AGAAGAGTCCCTGACTCTTGTGTGAAGTGTCTTCACAGACCGGAGG 2510
533 lLysLeuAspLeuLeuPheLeuArgGluGlyLeuLysAlaAlaLeuLys 549
2511 GATTAGATCTGCTGTTCTTAAGAAAGTGGGTTATGTGCACCTTAA 2560
550 GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSer 566
2561 GAAATAGTGTGCTTCTATGTAGATCACTCAGAGCCCTCAGAACATCCAT 2610
566 lAsnLysLeuArgLysLysLeuGluArgArgArgGluArgGluAla 583
2611 GAGCAACCTTAGAAGAGTTAGAGAGCGCTCAGAGGAAAGAGAGCTG 2660
583 spGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyrMetThr 599
2661 ACCAGGGGTGTTGAAGGATGTTCAACAGCTCTCTCTGATGAGACACC 2710
600 LeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 616
2711 CTGCTTCTCTCTGAGGGAGCCCTGATAGTCTCTCTCTCTCTCTCTCT 2760
616 rValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArg 633
2761 AGTTGGGCTTGTCTTAATTAATAGTGTGTGCTTGTGTTAGAGAGAG 2810
633 aISerAlaValGlnIleMetValLeuArgGlnAlaTyrGlnGlyLeu 649
2811 TGAGTCCAGTCCAGATCATGTGTTAGGCAACAGTACCAAGGCTTCTG 2860
650 SerGlnGlyLeuThrAspLeu 656
2861 AGCCAGAGAGAAACTACCTC 2881

seq.name: /cgn2/6/ptodata/2/ina/6A_COMB.seq:05-09-111-085-1
seq.documentation.block:
: Sequence 1, Application US/09111085
: Patent No. 6100034
: GENERAL INFORMATION:
: APPLICANT: Stoye, Jonathan P
: APPLICANT: Weiss, Robin A
: TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
: FILE REFERENCE: 4238/75168
: CURRENT FILING DATE: US/09/111.085
: EARLIER APPLICATION NUMBER: GB 9710154.7
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2462

```

```

;      TYPE: DNA
;      ORGANISM: Porcine retrovirus
US-09-111-085-1

```

```
alignment_scores:
  Quality: 2594.50      Length: 680
  Ratio: 4.420          Gaps: 11
Percent Similarity: 86.324  Percent Identity: 73.088
```

```
alignment_block:
```

Align seg 1/1 to: US-09-111-085-1 from: 1 to: 2462

[illegible]

905	GAGGCTCTGGGAGAAAGAAAGATCTCTTCTGACTATTGGCCGACAAATA	954
228	GIuThrgtGIuThrgtGIuProProVaLaMeGtGIuProAspLysVaLLeuL	244
955	GAACACTGATGAAACCCCGGGTTGCTATGACGACCAAAATAAGGGTTGGC	1004
244	aGIuGIuGIuProProLaLeuGIuProProHIsAsnLeuProVaLProG	261
1005	CGAACCAAGACCTCCCAATCCAGAACAG.....	1032
261	LnLeuThrSerLeuArgProAspLLehGInProProSerAsnSerThr	277
1033AGGCACTCTCCTTAACCCCTCTGATTCAATACAAACC	1068
278	ThrgtGIuLeuLeuProThrAsnThrProArgAsnSerProGIuVaLProVa	294
1069	TCGTGATCAGTCCCACT.....GAGCTCATCACTACT	1103
294	LIyThrgtGIuLnArgLeuPheSerLeuIleGInGIuLaPheGlaLaI	311
1104	TAAACACAGGGGAAACCTTTTACCCTCATCCAGAGGAGCTTTCAAGCTC	1153
311	LaAsnSerThAspProAspLaLthrSerSerCysTrpLeuCysLeuSer	327
1154	TTTACTCCACGACTCTCAAGAGCTACTCTTCTTGGCTTGGCTTAGCT	1203
328	SerGIuProProTyTrtGIuGIuMeLaLaLysGIuArgLysPheAsnVa	344
1204	TCGGGGCCACCTTACTATGAGGGAATGGCTAGAGGAGGAAATCAATGT	1253
344	ThrLysGIuHIsArgAsnGInCysThrTrpGIuSerArgAsnLysLeuTh	361
1254	GACAAAGGAACTTAAGACACAAATGACATGGGGATCCCAATAATGACTTA	1303
361	hTrLeuThrgtGIuVaLserGIuLysGIuThrCysIleGIuLysLaIProPro	377
1304	CCCTTACTAGGTTTCTGGAAAAGGCACTGCAATAGGAGATGGTCCCCCA	1353
378	SerHIsGIuHIsLeuCysTrtSerThrVaLaLTruGIuGInLaIaSerGI	394
1354	TCCACCAACACCTTGTTAACACACACTGAAACCTTTATCCGAAACCTCTGA	1403
394	uAsnGInTrtLeuVaLProGIuTrtYrAsnLaArgTrtTrpLaLeuAsnThrg	411
1404	GACTCATATCTCGTAACCTGGTTATGACAGAGGGGAGCAATGTAATCTG	1453
411	LIyLeuThrProCysVaLserThrSerVaLpheAsnGInSerLysAspPhe	427
1454	GATTAAACCCCTGTGTGTTTCCACCTTGTTGTTTCAACCAAACTAAAGACTTT	1503
428	CysVaLMeTrVaLGIuLnLeuLaProArgVaLTruTrtHIsProGIuGIuVa	444
1504	TGGGTTATGGTCCAAATGTGTCCCGGGGTGATACGTACGTCCGAAAAAGC	1553
444	IVaLLeuAspGIuTrtYrAspTrtArgTrtYrAsnArgProLysArgGIuProV	461
1554	AGTCTGTGATGAAATGACTATATGATATATATCGCCAAAAGAGAGGCCA	1603
461	aLserLeuThrLeuLaVaLMeLeuGIuLysGIuThrAlaVaLGIuVaL	477
1604	TATCTCTGACACTAGCTGTGATGCTCGGATTTGGAGTGGCTGCAAGCGCTG	1653
478	GIuThrGIuThrVaLaLaLeuLeuThrGIuProGInGInLeuGIuLysGI	494
1654	GGAACAGGAACGGCTCCCTCAATACAGAGACCGCAACAGCTGGAGAAAGG	1703
494	YLLeuGIuLeuLeuHIsAlaLaMeThrGIuAspLeuArgAlaLeuLysG	511
1704	ACTTAGTAACCTACATCGAATGTGTAAAGGAAATGTCCAAAGCCCTAGAA	1753
511	LIuSerVaLserAsnLeuGIuGIuLysLeuThrSerLeuSerGIuVaLaI	527
1754	AATGTGCATGAACTCGAGAGAAATCCCTAACTCTTATCTGAAGTGGTT	1803


```

528 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlyGlyLeu 544
1804 CTRACAGACAGAGAGGGGTTAGATCTGTTATTTCTAAAGAGAGGGTT 1853
544 UCysAlaAlaLeuLysGluGlyCysCysPheTyrValAspHisSerGly 561
1854 ATGTGAGCCTTAAAGAGAGATGCTGCTTATATAGATCACTCAGAG 1903
561 IallArgAspSerMetAsnLysLeuArgLysLysLeuGlnArgArg 577
1904 CCATCAGAGACTCCATGAGCAAGCTTAGAGAAAGTTAGAGAGGGCTGA 1953
578 ArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArg 594
1954 AGGGAAGAGAGGCTGACACAGGGGTGTTGAAAGATGTTCAACAGGCTC 2003
594 rProThrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValAla 611
2004 TCCTTGATGACACACCTGCTTCTGCTCTGACGGGGCCCTAGTAGTCC 2053
611 euleLeuLeuLeuThrValGlyProCysLeuLeuAsnArgPheValAla 627
2054 TGCTCCTGTTACTTACAGCTGCGCTTGTAAATTAATAGTTGTTGCC 2103
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGln 644
2104 TTGTGTAGAGAACGAGTGTGCTGCTGATGATGATGATGATGATGATG 2153
644 nTyrGlnGlyLeuLeuSerGlnGlyGlyLeuThrAspLeu 656
2154 GTACCAAGGCTCTGAGCCAGAGAACTGACCTC 2190
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-2
seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MCP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-2

alignment_scores:
    Quality: 2497.50      Length: 678
    Ratio: 4.328          Gaps: 13
    Percent Similarity: 85.103      Percent Identity: 71.829

alignment_block:
    US-09-171-553b-6 x US-08-766-528-2

Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333

4 ThrLeuSerArgArgHisLeuProThrArgGlyGlyLeuProLysArgLe 20
|||||
4757 ACCTTATCTGTTACTTTTGACACACACAGCGCTGTGAAAGTCA..... 4801
20 uArg.IleProLeuSerPheAlaSerIleAlaTrpPheLeuThr 36
|||||
4802 .AGGATCCCTTAAGCTTCCCTCCATCGCGGTCTTCTCTCTCA 4850
37 IleThrProGlnAlaSerSerLysArgLeuLeuAspSerSerAsnPro 53
|||||
4851 ATATACCTCAAGTAAAGTAAAGCTTGTGAGACAGCCGCACTCCCA 4900
53 sArgProLeuSerLeuThrTrpLeuIleLeuAspProAspThrGlyVal 70
|||||
4901 TTAACCTTATCTCTCCTGTTACTTACTTACTTACTTACTTACTTACT 4950
70 hrValAsnSerThrArgGlyValAlaProArgGlyThrTrpProGly 86
|||||
4951 ATATTAACAGCACTCAAGGAGGCTCCCTTGGGACCTGCTGCTGCA 5000
87 LeuHisPheCysLeuArgLeuLeuAsnProAlaVal.....LysSerTh 101
|||||
5001 TTATATGTCTGCTTGCATGATGATGATGATGATGATGATGATGATG 5050
101 rProProAsnLeuValArgSerTyrGlyPheTyrCysProGly...T 117
|||||
5051 ACCCCCGATGTAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 5100
117 hgGluLysGluLysTyrCysGlyGlySerGlyGlySerPheCysArg 133
|||||
5101 CAAATATGAAGAATATTTGTGAATCTCAGGATTTCTTTTGAAGCA 5150
134 TrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 150
|||||
5151 TGGAGCTGCATAACTTAAATGATGGAATGGAATGGAATGGAATGCA 5200
150 uGlnAspArgValLysPheSerPheValAsnSer..... 161
|||||
5201 GCAACACAGAGTAACTTCTTTTGTTAACAAATCCACGATTATATC 5250
162 .....GlyProGlyLysTyrLys.....MetMetLysLeuTyr 172
|||||
5251 AATTATATTATGAGGATGAGATGAGAAATGGAACAGCGGTCA 5300
173 LysAspLys.....SerCysSerProSerAspLeuAspTyr 184
|||||
5301 AAAGTGTACGAATTAAGCAAAATAGCTGTCATTCTTAACCTAGATTA 5350
184 rLeuLysIleSerPheThrGlu...ArgLysThrGlyLysTyrSerLysVa 200
|||||
5351 CTTAAATAATAGTTTCACTGAAAGAAAGCAACAGAAATTTCAAAAGT 5400
200 LAspLysTrpTyrGluLeuGlyAsnSer.PheLeuLeuTyrGlyGly 216
|||||
5401 GGTAAATGTATATCTTGGGAATAGTATAC.....TATGAGGCTCT 5444
217 Ala.....GlySerThrLeuThrIleArgLeuArgIleGluThrG 230
|||||

```

5445 GGGAGAAAAGAGATCTGTTCTGACATATTCGCCCTCAGAAATAGAACTCA 5494
 230 YTHG1UPROVALALAMEGLYPROASPLYSVALLEU1AGLUNG 247
 |||||||.....
 5495 GATGAACCTCGGTGCTATAGACCAATTAAGGTTTGCCGACACAG 5544
 247 LYPROVALLEU1UPROHISANLEUPROVALPROGLINLEUTHR 263
 |||||||.....
 5545 GACCTCAATCCAGAACAG..... 5564
 264 SERLEUARGPROASPLIETHRGLINPROPROSERASERHTRHGLYLE 280
 |||||||.....
 5565AGGCATCTCCTCAACCCCTGATTCAATCAACCTCTGATC 5608
 280 ULEPROTHASANTHRPROARGASERPROGLYVALPROVALYTHRG 297
 |||||||.....
 5609 AGTCCCACT.....GAGCCTAACATCATTATTAACAG 5643
 297 LYLINARGLEUPHESERLEU1LEGLINLYALAPHEGLINALEASNSER 313
 |||||||.....
 5644 GGGGAAACTTTTAGCCTATCCAGGAGCTTTCAAGCTTAACTCC 5693
 314 THRSPROASPLATHRSERSETCYSTRPLEUCYSEUSERSERGLYPT 330
 |||||||.....
 5694 ACGACTCCAGAGGCTACCTCTTGTGGCTTGTAGCTTCCGCGCC 5743
 330 OPROTYTRGLUGLYMETALALYSGLUARGLYSPHEASVALHTRYSG 347
 |||||||.....
 5744 ACCTTACTATAGGGAATGGCTAGAGAGGAAATTCATGTGCAACAG 5793
 347 LUHISARGASGLINCYSYTHRTYRGLYSERARGANLYSLEUTHRLEUTHR 363
 |||||||.....
 5794 AACATAGAGACCAATGTACATGGGATCCCAATTAAGCTTACCTTACT 5843
 364 GLUVALSERGLYSGLYTHRCYSLLEGLYLSALAPROPROSERHISGL 380
 |||||||.....
 5844 GAGTTTCTGGAAAAGGACCTGCATAGGATGGTCTCCCATCCACCA 5893
 380 NHISLEUCYSTYRSETHRVALVALTYRGLINLALASERGLUNASGLINT 397
 |||||||.....
 5894 ACACCTTTGTACACACACTGAAGCCTTAATCGAACCCTGAGAGTCAAT 5943
 397 YRLEUVALPROGLYTYRASNARTPTTPALACYASANTHGLYLEUTHR 413
 |||||||.....
 5944 ATCGGTACCTGTTATGACAGTGGGCGCATTAATGATGATTAAC 5993
 414 PROCYSVALSERHRSERVALPHEASGLINSELYSASPHCYSVIME 430
 |||||||.....
 5994 CTTGTGTTCCACCTTGTTTCAACCAAACTAAAGACTTTTGCTTAT 6043
 430 TVALGINLEVALPROARGVALTYRTYRHSISPROGLINLVALLEUA 447
 |||||||.....
 6044 GGTCAAAATTTCCCGGGGTGTACTATCCCGAANAAGACATCTCTTG 6093
 447 SPGLUYRASPTRYRGTYSANARTPROLYSARGLUPROVALSERLEU 463
 |||||||.....
 6094 ATGATATGACTATATATATATGCGCAAAAAGAGACCATATTCCTG 6143
 464 THRLEUVALMETLEUGLYLEUGLYTHRALVALGLYALGLYTHRGL 480
 |||||||.....
 6144 ACCTAGCTGTATATGCTCGATTGGAGTGGCTCAGCGCGGAGAACG 6193
 480 YTHRALALEU1LEUTHRGLYPROGLINLLEUGLYLUSGLYLEUGLYG 497
 |||||||.....
 6194 AACGGCTCCCTAATCACAGACCGCAACAGCTGAGAAAGACTTATGTA 6243
 497 LULEUHSIALALAMETHRGLUSPLEUARGALALEULYSGLUSERVAL 513
 |||||||.....
 6244 ACCTACATGCAATGTATGAGAGATCTCCAAAGCCATGAAGAAATCTGTC 6293
 514 SERANLEUGLUGLUSERLEUTHRSERLEUSERGLVALYALLEUGLINS 530
 |||||||.....
 6294 AGTAACCTGGAGGATCCCTAACCCTTATCTGAGGTGTTACAGAA 6343

530 NARGARGLYLEUASPLEU1LEUPHEUARGGLUGLYLEUCYSALAA 547
 |||||||.....
 6344 CAGAAGGGGTTATGATCTGTTATTTCTAAAGAGAGAGGTTATGTTAG 6393
 547 LALEULYSGLU1UCYSPHETRYVALASPHISERLYALALEARG 563
 |||||||.....
 6394 CTTTAAAGAGGAATGCTCTTATATGATGATCCTCAGACCAACACAA 6443
 564 ASPSERMETASNLYSLEUARGLYSLYSLEUGLUARGARGARGGLUAR 580
 |||||||.....
 6444 GACTCATGAGCAAGCTTAGAGAAAGTTAGAGAGGCTCGAAGGGAAG 6493
 580 GGLUALASPSGLINLYTRPHEGLUGLYTRPHEASNARGSERPROTRPM 597
 |||||||.....
 6494 AGAGGCTGACCAAGGGTGTGTAAGCATGTTCAACAGGTCTCTTGA 6543
 597 ELTHRTHRLEU1SERLALALEUTHRGLYPROLEUVALLEU1LEU 613
 |||||||.....
 6544 TGACCAACCTGCTTCTGCTGACCGGCGCCCTAGTACTGCTGCTCG 6593
 614 LEU1LEUTHRVALGLYPROCYSEU1LEASARGPHEVALALAPHEVAL 630
 |||||||.....
 6594 TTACTTACAGTTGGGCTTGTATTAATTAATGATTTGCTTGTGTTAG 6643
 630 GGLUARGVALSERLALALEU1LEU1LEUVALLEUARGGLINLYRGLN 647
 |||||||.....
 6644 AGAAGCAGTGAAGTGCAGTGCAGATCATGTGACTTAGCAACAGTACCAAG 6693
 647 LYLEU1LEUSERGLINLYGLUTHRSPLEU 656
 |||||||.....
 6694 GCCTTCTGAGCCAGAGAACTGACCTC 6722

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-766-528-1

seq_documentation_block:
 ; Sequence 1, Application US/08766528
 ; Patent No. 6190861
 ; GENERAL INFORMATION:
 ; APPLICANT: Jay A. Fishman
 ; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 ; NUMBER OF SEQUENCES: 74
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/766,528
 ; FILING DATE:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/572,645
 ; FILING DATE: 14-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Louis Myers
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: MCP-038CP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8060 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-1

alignment_scores:
Quality: 2360.00 Length: 656
Ratio: 4.252 Gaps: 11
Percent Similarity: 84.604 Percent Identity: 68.445

alignment_block:
US-09-171-553b-6 x US-08-766-528-1

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1 MethisProThrlleuSerArgArghisLeuProThrlleuArgglylgluPr 17
|||||
86 ATGCATCCCGCGTTAAGCGCGCCACCCCGATTCGGGGTGAAGGCC 135
|||||
17 ollysArgleuArglleProleuSerPheAlaSerlleAlaTrpPheLeu 34
|||||
136 GAAAGAGCTGAAATGCCCTTAAGCTTCGCCCTCCATCGCGTTCCTTA 185
|||||
34 hrleuThrlleThrlleProglAlaSerSerlyArgleuIleAspSer 50
|||||
186 CTCTGCATATACCTCTCAGACTATGGATGCCATAGAGACGCCCTG 235
|||||
51 AsnProHisArgProleuSerleuThrlleuIleAspProAspTh 67
|||||
236 AACGCCCAAAACCTTATCTCTCACCTGTTAATTACTGACTCGCGCAC 285
|||||
67 rglValThrlleThrlleThrlleThrlleThrlleThrlleThrlle 84
|||||
286 AGCTTATATATCAACAACACTCAAGGAGGAGCTCTTTAGGAACCTGT 335
|||||
84 rProgluLeuHisPheCysleuArgleuIleAsnProAlaValylSer 100
|||||
336 GCCCTGATCTATACCTTGGCCACATGATTTCTTACGTGACCTCA 385
|||||
101 ThrProAsnLeuValArgSerlyrlyPheThrlleCysCysProglyTh 117
|||||
386 ...CCCCAGATATCTCCATGCTCACGATTTATGTTGCCAGAGACC 432
|||||
117 rgluysglu...LysTrpCysgluLysSerlyrlygluSerPheCysArg 133
|||||
433 ACCAAATATGGAACATATGGGAAATCCAGAGATTTCTTTGTAAAC 482
|||||
133 rglTrpSerCysValThrSerAsnAspGlyAspTrpTrpProIleSer 149
|||||
483 AATGGAACCTGTACCTCTATATGATGATTTGGAATGGCCAAACCTCT 532
|||||
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProgl 164
|||||
533 CAGCGAGATAGGTAAGTTTCTTATGTCACACCTATACCACTCTGG 582
|||||
164 yLysTrpLysMetLysleuThrlleThrlleThrlleThrlleThrlle 179
|||||
583 ACAATTTATTAACCTGACCTGATGATTAAGAACTGGAAGCCCAAGGCTCTC 632
|||||
179 rOserAspLeuAspTrpLeuLysIleSerPheThr..gluArglyThr 195
|||||
633 CTTCAGACCTAGATTACTATAAATTAAGTTTCTTCTGAGGAAATGATAT 682
|||||
195 lylLysTrpSerLysValAspLysTrpTrpTrpTrpTrpTrpTrpTrp 211
|||||
683 GAAATATATCTTAATAATGATGATGATGATGATGATGATGATGAT 730
|||||
211 uLeuThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 225
|||||
731 ...TATGAGGCTGCGGTAAACAACAGAGCTCTTCACTTCACTATTCGCC 776
|||||
225 euArglleGluThrlleThrlleThrlleThrlleThrlleThrlleThrlle 241
|||||
777 TCAAAATTAACAG...CTGAGGCTCCCAATGCTATAGACCAAAATACG 823
|||||
242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
|||||
824 GTCTTGAAGGCTCAAGACCCCAACCAAGGACCA..... 859
|||||
258 oValProgluLeuThrlleThrlleThrlleThrlleThrlleThrlleThrlle 275
|||||
860 .....GACACATCTCTTA 872
|||||
275 snserThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 291
|||||
873 ACATTAACCTTGTGATCAACCCCACTGATGATGATGATGATGATGATGAT 916
|||||
292 ValProValLysThrlleThrlleThrlleThrlleThrlleThrlleThrlle 308
|||||
917 .....AAATGGGGCAAACTTTTATGCTCTATCCAGGAGGCTTT 957
|||||
308 eGlnAlaIleAsnSerThrlleThrlleThrlleThrlleThrlleThrlleThrlle 325
|||||
958 TCAAGCTCTTAACCTCAACGACCTCCAGAGCTTCTTCTTGTGGCTAT 1007
|||||
325 yLysLeuSerSerlyrlyProProThrlleThrlleThrlleThrlleThrlle 341
|||||
1008 GCTTATGCTTGGGCCCACTTATGATGAAGATGCTTAAGAGGAGGAAA 1057
|||||
342 PheAsnValThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 358
|||||
1058 TTCAATTTGACAAAGAACATAGAGACCATGACATGGGAGATCCCAAAA 1107
|||||
358 nLysLeuThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 375
|||||
1108 TAACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 1157
|||||
375 lAProProSerHisGlnHisLeuCysTrpSerThrlleThrlleThrlleThrlle 391
|||||
1158 TTCCCCCATCCCAACACCTTGTATGACACACTGAAGCTTATATCA 1207
|||||
392 AlAserGluAsnGlnThrlleThrlleThrlleThrlleThrlleThrlleThrlle 408
|||||
1208 ACCTGTAAGTCAATATCTGTATGATGATGATGATGATGATGATGATGAT 1257
|||||
408 sAsnThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 425
|||||
1258 TAATACGATTAACCCCTTGTGTTCCACTTGTGTTTAAACCAAACTA 1307
|||||
425 yAspPheCysValMetValGlnIleValProArgValThrlleThrlleThrlle 441
|||||
1308 AAGATTTTTCATTAATGATGATGATGATGATGATGATGATGATGATGAT 1357
|||||
442 GluGluValValLeuAspGluThrlleThrlleThrlleThrlleThrlleThrlleThrlle 458
|||||
1358 GAAAGAACATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1407
|||||
458 gLiuProValSerleuThrlleThrlleThrlleThrlleThrlleThrlleThrlle 475
|||||
1408 AGAACCATATCTCTGACACTTCTGATGATGATGATGATGATGATGATGAT 1457
|||||
475 alGlyValGlyThrlleThrlleThrlleThrlleThrlleThrlleThrlleThrlle 491
|||||
1458 CAGGTGATGACAGACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
|||||
492 GluLysGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGly 508
|||||
1508 GAACAGAGACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1557
|||||
508 alLeuLysGluSerValSerAsnLeuGluGluSerleuThrlleThrlleThrlle 525
|||||
1558 CCTAGAGAAATCTGTCACTGATGATGATGATGATGATGATGATGATGATGAT 1607
|||||
525 luvAlaValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlu 541
|||||
1608 AAGTATCTCTACAGATTAAGAGGTTGATTTATTTATTTCTAAAGAA 1657

```

```

542 GYGLYLeucysAlaAlaLeuLysGluGluCysCysPheTyrValAspH 558
|||||.....
1658 GGAGGATTATGTACCTTGAAGGAGGATGCTTTTATGTGATCA 1707
|||||.....
558 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGluA 575
|||||.....
1708 TTCAGGGCCATCAGAGCTCCATGAAACAACTTAGAGAAAGTTGGAGA 1757
|||||.....
575 rGaArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPhe 591
|||||.....
1758 AGCGTCGAAGGAGAAAGAACTACTCAGAGGTGTTGAGGGATGCTTC 1807
|||||.....
592 AsnArgSerProTyrMetThrThrLeuSerAlaLeuThrGlyProLe 608
|||||.....
1808 AACAGGTCTCTGTGGTGGCTACCTTCTTCTTTACAGGACCTT 1857
|||||.....
608 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArg 625
|||||.....
1858 AATAGTCTCTCTGTTACTACAGATTGGCCATGATATTATTAACAAGT 1907
|||||.....
625 heValAlaIleValAlaArgValSerAlaValGlnIleMetValLeu 641
|||||.....
1908 TAAATGGCTTCATTAGAGAACGAATAGTGCAGATCATCTGTTACT 1957
|||||.....
642 ArgGlnGlnTyrGln 646
|||||.....
1958 AGACACAGTACCA 1972
|||||.....

```

```

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-766-528-3
seq_documentation block:
; Sequence 3, Application US/08766528
; Patent No. 6190861

```

```

; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-3

```

```

alignment_scores:
    Quality: 2353.00    Length: 656
    Ratio: 4.232        Gaps: 12
    Percent Similarity: 84.756    Percent Identity: 68.598

```

```

alignment_block:
US-09-171-553b-6 x US-08-766-528-3

```

```

Align seg 1/1 to: US-08-766-528-3 from: 1 to: 8132

```

```

1 MethHisProThrLeuSerArgGlnHisLeuProThrArgGlyGluPr 17
|||||.....
5620 ATGCATGCCACGCTTAAACCGGCCGCCACCTCCGATTCGGGGTGAAGCC 5669
|||||.....
17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTyrPheLeuT 34
|||||.....
5670 GAAAGACGTGAATCCCTTAAGCTTCGCTCCATCGGTGGTCTCTTA 5719
|||||.....
34 hIleuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
|||||.....
5720 CTCTGTCATTAACCTCTCAGACTAATGATGGCATAGGAGACACCTG 5769
|||||.....
51 AsnProHisArgProLeuSerLeuThrTyrPheIleIleAspProAspH 67
|||||.....
5770 AACTCCCATTAACCTTATCTCTCACCTGGTTAATTAACGCTCCGCGAC 5819
|||||.....
67 rGlyValAlaThrValAsnSerThrArgGlyValAlaProArgGlyThrTyr 84
|||||.....
5820 AGGTATTAAATATCAACAACCTCAAGGAGGAGGCTCTTTAGGAACTGGT 5869
|||||.....
84 rProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||.....
5870 GCGCTGATGTCATACGTTGGCTCAGATCAGTTTCTGATGTCGACTCA 5919
|||||.....
101 ThrProAsnLeuValAlaArgSerTyrGlyPheTyrCysCysProGlyTh 117
|||||.....
5920 ....CCCCAGATATCCTCCATGCTCAGCGATTTATGTTGGCCAGAGCC 5966
|||||.....
117 rGluLysGlu...LysTyrCysGlyGlySerGlyGlySerPheCysArg 133
|||||.....
5967 ACCAATTAATGAAACATTTGCGGAATCCAGAGATTTCTTTTGAAC 6016
|||||.....
133 rGTrpSerCysValThrSerAsnAspGlyAspTyrPyrTrpProLysSer 149
|||||.....
6017 AATGGAACGTGTAACTCTAATGATGATGGAATGGAATGCCAACCTCT 6066
|||||.....
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
|||||.....
6067 CAGCAGGATAGGGTAAAGTTTCTTATGTCACACCTATACGACGCTCTG 6116
|||||.....
164 yLysTyrLysMetIleLysLeuTyrLysAspLysSer.....CysSerP 179
|||||.....
6117 ACAATTTAATTAACCTGACTGATTAAGACTGGAAGCCCAAGTGTCTTC 6166
|||||.....
179 rSerAspLeuAspTyrLeuLysIleSerPheThr...GluArgLysThrG 195
|||||.....
6167 CTTCAGACCTAGATTACCTTAAATAAATAGTTTCACTGGAAGGAAAAACA 6216
|||||.....
195 yLysTyrSerLysValAspLysTyrTyrGluLeuGlyAsnSer...PheLe 211
|||||.....
6217 GAAATATCTCTAAATGAGTAAATGATGTCTTGGGGAATGGTAAAT... 6264
|||||.....
211 uLeuTyrGlyGlyValAla.....GlySerThrLeuThrIleArgL 225
|||||.....
6265 ....TATGAGGCTCGGGTAAACAACAGGCTCATTTCTACTATATTCGCC 6310
|||||.....
225 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 241
|||||.....
6311 TCAAAATTAACCAAG...CTGAGGCTCCCAATGCTATAGAGCAAAATAG 6357
|||||.....
242 ValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 258
|||||.....

```

```

6358 GTCTTGACGGGTCAAGACCCCAACCAAGACCA..... 6393
258 OVALPROGInleuthrSerleuArProspilEthrGInProProSera 275
6394 .....GGACCATCCTTA 6406
275 snSerThrThrGlyLeuIleProthrAsnThrProArGsnSerProGly 291
|| ||||| ||||| |||||
6407 ACATTAATCTTGATCAGACCCCACTGAGTCT.....AACAGC..... 6444
292 ValProValysThrGlyGlnArgLeuPheSerleuIleGlnGlyAlaPh 308
||||| ||||| ||||| |||||
6445 ...ACGACTAAATGGGGCAAAACTTTTACCTCATCCAGAGGCTTT 6491
308 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeu 325
||||| ||||| ||||| |||||
6492 TCAGTCTTAACTCCAGCTCCAGAGCTCCTCTCTCTGTTGCTAT 6541
325 yslSerSerGlyProProTyrTyrGlnGlyMetAlaGlyGlnArg 341
||||| ||||| ||||| |||||
6542 GCTTACCTCGGGCCACCTTCTATGAAGAAATGGCTAGAAGAGGAAA 6591
342 PheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySerArg 358
||||| ||||| ||||| |||||
6592 TTCATGTGACAAAGAACATAGACCAATGCATGGGGATCCCAAA 6641
358 nlySerLeuThrLeuThrGlnValSerGlyLysGlyCysIleGlyLys 375
||||| ||||| ||||| |||||
6642 TAAAGCTTACCTTACTGAGGTTTCGAAAGACCTGCATAGCAAAAG 6691
375 lApProSerHisGlnHisLeuGlyCysTyrSerThrValValTyrGln 391
||||| ||||| ||||| |||||
6692 TTCGCCCATCCCAACACCTTTTAAACACACCTGAAGCCCTTATATA 6741
392 AlaSerGlnAsnGlnTyrLeuValProGlyTyrAsnArgTrpTyrAla 408
||||| ||||| ||||| |||||
6742 ACCTGTGAGAGTCATATCTGTGCTGATTAAGACAGTGGTGCA 6791
408 sAspThrGlyLeuThrProGlyValSerThrSerValPheAsnGln 425
||||| ||||| ||||| |||||
6792 TAAATCTGATTAACCTTGCTGTTCCACCTTGCTTTTAAACAACTA 6841
425 yAspPheCysValMetValGlnIleValProArgValTyrTyrHis 441
||||| ||||| ||||| |||||
6842 AAGATTTTGCATATGTGTCAAATTTGTTCCCGAGGTATTAATACCC 6891
442 GlnGlnValValLeuAspGlnTyrAspTyrArgTyrAsnArgProLys 458
||||| ||||| ||||| |||||
6892 GAAAGACATCTGATGATATGACTACAGAAATCATCAACAAAGAG 6941
458 GlnProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAla 475
||||| ||||| ||||| |||||
6942 AGAACCCATATCTGACACTGCTGTGATGCTCGAGCTTGAGTGCAG 6991
475 alGlyValGlyThrGlyThrAlaIleuIleThrGlnGlyProGlnGln 491
||||| ||||| ||||| |||||
6992 CAGGTGTAGGACAGGACAGCTCCCTGGTCACGGGACCAACAGCTA 7041
492 GlnLysGlyLeuGlyGlnLeuHisAlaIleMetThrGlnLysPheVal 508
||||| ||||| ||||| |||||
7042 GAACAGAGACTTAGTAACTCATCGAATTTCTAACAGAGATCTCCAGC 7091
508 alLeuLysGlnSerValSerAsnLeuGlnGlnSerLeuThrSerLeu 525
||||| ||||| ||||| |||||
7092 CCTAGAAAATCTCTCAGTAACCTGAGGAATCCCTAACCTTATCTG 7141
525 lValValleuGlnAsnArgArgGlyLeuAspLeuPheLeuArgGln 541
||||| ||||| ||||| |||||
7142 AAGTAGTCTTACAGATAGAGAGGTTAGATTATTTCTAAAAAGAA 7191
542 GlyLysLeuCysAlaIleLeuLysGlnGlyCysCysPheTyrValAsp 558
||||| ||||| ||||| |||||
7192 GGAGATTATGTGTAGCTTGAGAGAGGAATGCTTTTATGTGATCA 7241

```

```

558 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuGln 575
||||| ||||| ||||| |||||
7242 TTCAGAGGCCCATCAGAGACTCCATGAACAACTAGAGAAAGTTGGAGA 7291
575 rGArgArgArgGlnArgGlnAlaAspGlnGlyTrpPheGlyTrpPhe 591
||||| ||||| ||||| |||||
7292 AGCGTCGAAGGAAAGAAACTACTCAAGGTTGTTGAGGATGCTTC 7341
592 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPro 608
||||| ||||| ||||| |||||
7342 AACAGGCTCTTGTGTTGGCTACCTTCTTCTTAAACAGAGCCCTT 7391
608 uValValleuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArg 625
||||| ||||| ||||| |||||
7392 AATAGTCTCTCTCTGTTACTACAGATTGGGCATGTATATTATACAG 7441
625 heValAlaIlePheValArgGlnArgValSerAlaValGlnIleMetVal 641
||||| ||||| ||||| |||||
7442 TAAATGCTTTCATTAAGAACGAATTAAGTCCAGTCCAGATCATGTCT 7491
642 ArgGlnGlnTyrGln 646
||||| ||||| ||||| |||||
7492 AGACAAACGATACCA 7506

seq_name: /cgn2_6/pdata/2/lna/6B.COMB.seq:US-09-376-781-25
seq_documentation_block:
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

alignment_scores:
Quality: 2136.00 Length: 654
Ratio: 3.963 Gaps: 14
Percent Similarity: 82.416 Percent Identity: 64.526

alignment_block:
US-09-171-553b-6 x US-09-376-781-25 ..

Align seg 1/1 to: US-09-376-781-25 from: 1 to: 2000

1 MetHisProThrLeuSerArgArgHisLeuProThrArgGlyGlyGln 17
||||| ||||| ||||| |||||
43 ATACATCCACGCTTAAAGCCGCGCCACCTCGATTCAGAGGTGAAGACC 92
17 olysArgLeuArgGlyLeuProLeuSerPheAlaSerIleLeuArgPheLeu 34
||||| ||||| ||||| |||||
93 AAAAAGACTAAATCCCTTAAAGCTTGGCTCCATCAACATGCTTCTTA 142
34 hrLeuThrThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50

```

```

|||||..... |||.....: |||..... |||.....
143 CTCTGTCATTAACCTCTCAGACTAATGATGACATAGAGACACCTTG 192
|||||..... |||.....: |||..... |||.....
51 AsnProHisArgProLeuSerLeuThrTrpLeuIleLeuAspProSph 67
|||||..... |||.....: |||..... |||.....
193 AACCCCATTAACCTTATCTCTGACCTGTTAATTACTGACTCTGACAC 242
|||||..... |||.....: |||..... |||.....
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||..... |||.....: |||..... |||.....
243 AGGATTAATATCCACAGCGCTGAGGAGGAGGCTCCTTAGAAACCTGGT 292
|||||..... |||.....: |||..... |||.....
84 rProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||..... |||.....: |||..... |||.....
293 GGCCTGATCTATATGCTGCTCAGATCATCTCTAGCTGACCTCA 342
|||||..... |||.....: |||..... |||.....
101 ThrProAsnLeuValArgSerThrGlyPheThrCysCysProGlyThr 117
|||||..... |||.....: |||..... |||.....
343 ACC...CCAGATATCTCCGCTTACGATTTTATGTTGCCAGGAC 389
|||||..... |||.....: |||..... |||.....
117 rGluLysGlu...LysThrCysGlyLysSerGlyLysSerPheCysArg 133
|||||..... |||.....: |||..... |||.....
390 ACCAAATATAGAAACACATATGGAATCTAGAGATTCTTTACAAAC 439
|||||..... |||.....: |||..... |||.....
133 rGTPSerCysValThrSerAsnAspLysArgTrpLysTrpProIleSer 149
|||||..... |||.....: |||..... |||.....
440 AATGAGAGCTGTGTAACTCTATGATGAAATCGAAATGCGCAACCTCT 489
|||||..... |||.....: |||..... |||.....
150 LeuGlnAspArgValLysPheSerPheValAsnSerGlyProGlyLys 166
|||||..... |||.....: |||..... |||.....
490 CTGACGGATAGGGTAAAGCTTTTCTTATGTCAC...CCCATTAACA 533
|||||..... |||.....: |||..... |||.....
166 rLysMetLysLeu...TyrLysAspLysSer...CysSerProSerA 181
|||||..... |||.....: |||..... |||.....
534 CCGGACCTGAAACAGTCATACAGTACCATTTCTGGGTTTCCCTCAG 583
|||||..... |||.....: |||..... |||.....
181 sPLeuAspTyrLeuLysIleSerPheThrGlu...ArgLysThrLys 197
|||||..... |||.....: |||..... |||.....
584 ACCTAGATATCTTAAATAAGTTTCACCGAAATAAAGCAAGAAAT 633
|||||..... |||.....: |||..... |||.....
197 rYSerLysValAspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyr 213
|||||..... |||.....: |||..... |||.....
634 ATCCTAATAATGATTAATGATATGCTCGGGGAAATTAAT...ATATTATAC 680
|||||..... |||.....: |||..... |||.....
214...GlyLysGly...AlaGlySerThrLeuThrIleArgLeuArgIleG 228
|||||..... |||.....: |||..... |||.....
681 AGGTGGGACAGACAAACAGGCTCATCTTAACCATCGACTTAATAA 729
|||||..... |||.....: |||..... |||.....
228 LuThrGlyThrGluProProValAlaMetGlyProAspLysValLeuAla 244
|||||..... |||.....: |||..... |||.....
730...AGCCACGTAGAGCTCCCAATGCTATAGAGCGCAATAGCGTCTTAACG 777
|||||..... |||.....: |||..... |||.....
245 GluGlnGlyProProAlaLeuGluProProHisAsnLeuProValProG 261
|||||..... |||.....: |||..... |||.....
778 GGTCAAAGAACCCCAACCCAGGACCATCTCTGATATA... 816
|||||..... |||.....: |||..... |||.....
261 nLeuThrSerLeuArgProAspLysIleThrGluProProSerAsnSerThr 278
|||||..... |||.....: |||..... |||.....
817...ACTTCTAATTAGACCCCACTGAG...TCTAACACACAGA 853
|||||..... |||.....: |||..... |||.....
278 hGlyLeuIleProThrAsnThrProArgAsnSerProGlyValProVal 294
|||||..... |||.....: |||..... |||.....
854 CT... 855
|||||..... |||.....: |||..... |||.....
295 LysThrGlyLysArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI 311
|||||..... |||.....: |||..... |||.....
856 AAAAGGGGACAAACTTTTAGTCTCATCCAGGAGCTTTCAAGCTCC 905
|||||..... |||.....: |||..... |||.....
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerS 328
|||||..... |||.....: |||..... |||.....
906 TTACTCCAGACTCCAGAGCTACCTCTTCTGTTGGCTTCTTAACCT 955
|||||..... |||.....: |||..... |||.....
328 eGlyProProTyrTyrGlnGlyMetAlaLysGluArgLysPheAsnVal 344
|||||..... |||.....: |||..... |||.....
956 CGGGCCACCTTACTATAAGAAATGCGTAAAGAGAAAAATTCAATGTG 1005
|||||..... |||.....: |||..... |||.....
345 ThrLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuTh 361
|||||..... |||.....: |||..... |||.....
1006 ACAAAAAAATAGAGAACCAATGATAGGGATGATCCCAAAATAGCTTAC 1055
|||||..... |||.....: |||..... |||.....
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValaProProS 378
|||||..... |||.....: |||..... |||.....
1056 CCTTACTAGAGTTTCTCGAAAGACACCTGCAATAAAAAGTTCCCCCAT 1105
|||||..... |||.....: |||..... |||.....
378 eHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGlu 394
|||||..... |||.....: |||..... |||.....
1106 CCACCAACACCTTTTACACACACACTGAGCTTAAATCAACCTCTGAG 1155
|||||..... |||.....: |||..... |||.....
395 AsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrCl 411
|||||..... |||.....: |||..... |||.....
1156 AGTCATATCTGTGCTGCTGTTATGACAGGTGGGCAATGAAATACG 1205
|||||..... |||.....: |||..... |||.....
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheC 428
|||||..... |||.....: |||..... |||.....
1206 ATTAACCCCTTGTTGTTCCACCTTGTTTCAACCAACTAAAGACTTT 1255
|||||..... |||.....: |||..... |||.....
428 yValMetValGlnIleValProArgValTyrTyrHisProGluGluVal 444
|||||..... |||.....: |||..... |||.....
1256 ACATTAATGGTCCAAATTTGTTCCCGAGTATATTACTATCCAAAGAAACA 1305
|||||..... |||.....: |||..... |||.....
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
|||||..... |||.....: |||..... |||.....
1306 ATTCTGATGATATATGATTACAGAACCATCGCAAAAGAAACCAAT 1355
|||||..... |||.....: |||..... |||.....
461 lSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 478
|||||..... |||.....: |||..... |||.....
1356 ATCCCTGACACTCGCAGTAATGCTCGGACTCGGAGATGAACAGGTGTA 1405
|||||..... |||.....: |||..... |||.....
478 LysThrGlyThrAlaAlaLeuIleThrGlyProGlnLeuGluLysGly 494
|||||..... |||.....: |||..... |||.....
1406 GAACAGAACTGCAGCTTGTAGTTACAGACCTCAGAGCTAGAAACAGA 1455
|||||..... |||.....: |||..... |||.....
495 LeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
|||||..... |||.....: |||..... |||.....
1456 CTTAGTACCTTACATCAATTTGTAACGGAATCTCCAAACCTTAATAA 1505
|||||..... |||.....: |||..... |||.....
511 uSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 528
|||||..... |||.....: |||..... |||.....
1506 ATCTGTCAGTACCTGGAATAATCCATACCTCTTATCTGAGTAGTTC 1555
|||||..... |||.....: |||..... |||.....
528 eGluAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeu 544
|||||..... |||.....: |||..... |||.....
1556 TACAGATTAATAAAGGGTTAGATTATTTATTTCTAAATAAAGAGATT 1605
|||||..... |||.....: |||..... |||.....
545 CysAlaAlaLeuLysGluGluCysPheTyrValAspHisSerGlyAla 561
|||||..... |||.....: |||..... |||.....
1606 TGTGTACCTTAAAGAGAAATGCTGTTTATAGTATGATATTCAGGGGC 1655
|||||..... |||.....: |||..... |||.....
561 aLeuArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArg 578
|||||..... |||.....: |||..... |||.....
1656 CATGAGAGACTCCCATGAACAAGCTTAATAAAGGTGAGAGAAACGTGAA 1705
|||||..... |||.....: |||..... |||.....
578 rGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSe 594
|||||..... |||.....: |||..... |||.....
1706 GGGAAAGAGAACTTACTCAAGATGTTTAAAGAGTGTCAACAGGTC 1755
|||||..... |||.....: |||..... |||.....
594 rProThrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValVal 611
|||||..... |||.....: |||..... |||.....
1756 TCTTGTGTTGCTACCTTACTTCTTACTTTAAGAGACCTTATATATCC 1805
|||||..... |||.....: |||..... |||.....
611 eLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
|||||..... |||.....: |||..... |||.....
1806 TCTTCCTGTTACTCACAGATTGGCCATGATATTATTAACAAGTAAATGCC 1855
|||||..... |||.....: |||..... |||.....
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnI 644
|||||..... |||.....: |||..... |||.....
1856 TTCAATTAGAAACGAATAGTCCAGATCCAGATCATGTGTTACTTAGACACA 1905

```

100

```

1206 ATTAACCCCTTGTTCACCTTGTTTTCACCAACTAAAGCTTTT 1255
428 ysaVmetValGlnIleValProArGValTyrTyrHisProGluGluVal 444
1256 ACATTATGTCCTCAATTTGTTCCCGAGTATTTACTATCCCAAGAAACA 1305
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProVa 461
1306 ATTCCTGATGAATATGATTACAGGAACCATCGACAAACCAAAACCCAT 1355
461 LserLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValG 478
1356 ATCCCTGACACTGCGAGTATGCTCGACTCGAGAGTAAACAGCTGTCA 1405
478 LThrGlyThrAlaAlaLeuIleThrGlyProGlnGluLeuGlyGly 494
1406 GAACAGAACTGCGACTTGTAGTACAGAGCTCAGACAGCTAGAAACAGA 1455
495 LeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
1456 CTAGTACCTTACATCAAAATGTTAAACAGAAATCTCCAAAGCCCTAAA 1505
511 uSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 528
1506 ATCTGTCAGTACCTGGAATAATCCCTAACCTCTTATCTGAAGTATGTC 1555
528 euGlnAsnArgArgGlyLeuAspLeuPheLeuArgGluGlyLeu 544
1556 TACAGATATAAAAGGTTAGATTATTTCTTAAATAAAAGAAAGATTA 1605
545 CysAlaAlaLeuLysGluGlyCysPheTyrValAspHisSerGlyAl 561
1606 TGTGTACCTTAAAGGAAATGCTGTTTATGTAGATCATTCACAGGGC 1655
561 AlaArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgA 578
1656 CATCAGAGACTCCATGACAGCTTAAATAAAAGTTGGAAACGTCGAA 1705
578 rGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgS 594
1706 GGGAAAGAAACTTACTCAAGATGTTTAAAGATGGTTCAACAGCTC 1755
594 rProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValVal 611
1756 TCCCTGGTGGTACCTCTTCTTCTTAAACAGAGCCCTTAATATGTC 1805
611 euLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 627
1806 TCCCTCGTGTACACAGTTGGGCGCATTTATTAACAGTTAATTGCC 1855
628 PheValArgGluArgValSerAlaValGlnIleMetValLeuArgGln 644
1856 TTCATTAGAAACGAATAGTGCAGTCCAGATCATGCTACTAGACACA 1905
644 nTyrGln 646
1906 GTACCAA 1912
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-09-376-781-24
seq_documentation_block:
; Sequence 24, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Anderson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18

```

```

; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24

alignment_scores:
    Quality: 1627.00      Length: 491
    Ratio: 3.988          Gaps: 11
    Percent Similarity: 83.096      Percent Identity: 66.395

alignment_block:
US-09-171-553B-6 x US-09-376-781-24  ..

Align seg 1/1 to: US-09-376-781-24 from: 1 to: 1493

163 ProGlyLysTyrLysMetMetLysLeu...TyrLysAspLysSer...Cy 177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 CCCATTAACCAACCGGACCTGGAAACGTCATACAGATACAGTACATTCGGGCTG 53
177 sSerProSerAspLeuAspTyrLeuLysIleSerPheThrGlu...Argly 193
|-|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 TTTTCCCTCAGACCTAGATTATCTTAAATTAAGTTTCAACGCAAAAAA 103
193 sThrGlyLysTyrSerLysValAspLysTyrTyrGluLeuGlyAsnSerP 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 AACCAAGAAATATCTTAAATGATTAATGCTATGCTCGGGGAATATF. 152
210 heLeuLeuTyr.GlyGlyGly....AlaGlySerThrLeuThrIleAr 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 ..ATATTATACAGGTTGGAGACAGACACAGGCTCCATTTACCATCCG 200
224 gluLeuArgIleGluThrGlyThrGluProProValAlaMetGlyProAspL 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 ACTTAAATAA...AGCCAGCTAGAGCTTCATATGCTATAGACCGAATA 247
241 ysaValLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeu 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 CGGCTTACGGGCTCAAGAACCCACCCAGGACCATCTCTATATA 297
258 ProValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSe 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 .....ACTTCTAAATATAGACCCACGACTGAG.....TC 323
274 rAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProg 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 TAACAGCAGACT..... 336
291 LysValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAla 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 .....AAACGGGGACAAACCTTTTATGCTCATCCAGGAGCT 375
308 PheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLe 324
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 TTTCAGAGCTTAACTCCAGCAGCTCAGAGGCTACTCTTCTGTGGCT 425
324 uCysLeuSerSerGlyProProTyrTyrGluGlyMetAlaLysGluArgL 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
426 TTGCTTAACCTCGGCCCACTTACTATTAAGAAATGGCTAAAGAGAA 475
341 yspheAsnValThrLysGlnHisArgAsnGlyCysThrThrProLysArg 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 AATTCATGTGACAAACAAACATAGAGCAATGTACATGGGGATCCCAA 525
358 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLys 374

```



```

|||||
526 AATAAGCTTACCTTACTGAGCTTGTGAGAAAGACACCTGATATAAAAA 575
374 salapropioSerHisGlnHisLeucystySerThrValValItyGluG 391
576 GGTTCCTCCCATCCACACACCTTTACACACCACTGAAGCTTTATATC 625
391 lnaIserylGlnGlnItyrLeuValItyrGlyItyrAsnArgItyrPala 407
626 AAACCTCTGAGAGTCAATATCTGTACTGTTATGACAGTGGTGGGCA 675
408 CysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSe 424
676 TGTAAATCTGAGATTACCCCTGTGTCTCCACTGGTTTCAACCAAC 725
424 rItyAspPheCysValMetValGlnIleValProArgValItyrItyrHisP 441
726 TAAAGACTTTTACATTATGTGTCCAATATGTTCCCGAGATATTTACTATC 775
441 roGluGluValValLeuAspGluItyrAspItyrArgItyrAsnArgProlys 457
776 CCAGAAACCAATCTCGATGATATGATTACAGAACCATCGACCAAAAG 825
458 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAl 474
826 AAAAAACCATATCCCTGACACTCGCAGTAATGCTCGAGACTCGAGTAT 875
474 aValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGln 491
876 AACAGGTGTGAGACAGACAGACGAGCTTTAGTTACAGACTCGACGAC 925
491 euGluGlyGlyLeuGlyLeuHisAlaIleMetThrGluAspLeuArg 507
926 TAGAAACAGAGATAGTACCTACCTCAATGTAATGACAGAAATCTCCAA 975
508 AlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 524
976 GCCCTAAAAAATCTGTGCTAGTACTGAAAAATCCCTAACCTCTTATC 1025
524 rGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeuArg 541
1026 TGAAGTACTGTTACAGAAATAAAAAGGTTAGATTATTATTTCTAAAAA 1075
541 lueGlyLeuGlyCysAlaAlaLeuLysGluGlyCysPheItyrValAsp 557
1076 AAAGAGATTTATGTGTAACCTTAAAGGAGAAATGCTGTTTATGTAGAT 1125
558 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 574
1126 CATTGAGGGGCCATCAGAGACTCATGACAGCTTAAAAAAAGCTTGGA 1175
574 uArgArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyr 590
1176 GAAAGCTGAGGAGGAGAAAGAACTTACTCAAAAGATGTTTAAAGATGG 1225
591 PheAsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
1226 TTCAACAGGTCTCTTGTGTGGTACCTTACTTCTTAAACAGGAC 1275
1276 CTTAATAGTCTCTCTCTGTTACTACAGAGTTGGGCCATGATATATTACA 1325
607 OleuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsn 624
|||||
624 rgrPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 640
1326 AGTTAATGCTTCATTAGAAAGCAATAGTGTGAGTGCAGATCATGTGTA 1375
641 LeuArgGlnGlnItyrGln 646
|||||
1376 CTTAGACACAGTACCA 1393

```

seq_name: /cgn2_6/ptodata/2/lna/6B.COMB.seq:US-09-011-745-7

```

seq_documentation_block:
; Sequence 7, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 6312
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4058)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4061)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4060)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4059)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4247)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4246)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4248)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4249)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-7

alignment_scores:
Quality: 1508.50 Length: 746
Ratio: 3.060 Gaps: 24
Percent Similarity: 66.086 Percent Identity: 44.102

alignment_block:
US-09-171-553b-6 x US-09-011-745-7 ..
Align seg 1/1 to: US-09-011-745-7 from: 1 to: 6312

2 HisPro.....ThrLeuSerArgArgHisLeuProth 12
|||||
1428 CACCCGTTCCAGGTGCGGGGATCAAGTGTGTGTCAGAGCGCATCGACCCAG 1477

```

12 r.....ArglygluProLysArgLeuArgle 23
 : : : : :
 1478 CAGCCTGAGCCTCGTGGAAGGCCATACCTGCTGCTACATACC 1527
 23 roleSerPhe.....AlaSerIleAlaTrpPhe.leuThrleuth 36
 : : : : :
 1528 CGACCGCGGTAAAGATGATGATGCTGCTGCGTCCATGCTCTCAC 1577
 36 r.....IleThrProGlnAlaSerSerI 44
 : : : : :
 1578 CTCAAACCTGCACCCTTCGCGACAGATGAGTCTT...GGGAGCTGGA 1624
 44 ysArgLeuIle..... 47
 : : : : :
 1625 AAGAGCTGATCATCCTCTTAAGCTGCTATTGCGCGCGCGGAGCAGT 1674
 48 ...AspSerSerAsnProHisArgProLeuSerLeuThrTrpleuIle1 63
 : : : : :
 1675 CTGCAAAATTAAGAACCCCGACAGCCCATGACCTCTGCGAGTACT 1724
 63 eaSPProAspThrGlyValThrValAsnSerThrArgGlyValAlaProa 80
 : : : : :
 1725 GTCC...CAAACTGGAGACGTTGCTGGATACAAAGGCGAGTCCAGCCCC 1771
 80 rglGlyThrTrpProGluLeu.....HisPheCysLeuArgLeu... 93
 : : : : :
 1772 CTGGACTTGTGTGCCACACTTAACCTGATGATGCTGCTGGCGGCT 1821
 94IleAsnProAlaVally 99
 : : : : :
 1822 AGTCTGAGTCTGGGATATCCCGGAACCGATGCTGCTCTTAAGC 1871
 99 sSerThrProProAsn..... 104
 : : : : :
 1872 AGTCAGACCTCGGACTCAGACTACTGCGCTTATTAACAATCAGCT 1921
 105LeuValArgSer 108
 : : : : :
 1922 GGGAGCGCATAGGCTGACCTACCCTCGGCTAGGACTAGAAATGGCAAGC 1971
 109 TyrGlyPheTyrCysCysPro.....GlyThrGluLysGlu 121
 : : : : :
 1972 TCACCTTACGATGTCCCGGATGGCGGACCCCTTCAGAGCTGA 2021
 121 sTyrCysGlyGlySerGlyLeuSerPheCysArgArgTrpSerCysValT 138
 : : : : :
 2022 AAGGTGGGGGCTAGAAATCCCTATAGTAAAGATGGGATGTGAGA 2071
 138 hrSerAsnAspGlyAspTrpLysTrpProLleSerLeuGlnAspArgVal 154
 : : : : :
 2072 CCACGGGACCGGTTATTGG...CTATCTAATCTCCAAAGACCTCATR 2118
 155 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetLysIle 171
 : : : : :
 2119 ACGTAAATATGGGACCAAAATAGCAGATGACCTCAAAATTTCAACAGG 2168
 171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIle 188
 : : : : :
 2169 TCACCAAGCCGGCTGTTGTAACCC.....CTTAAATATG 2203
 188 erPheThrGluArgLysThrGlyLysTyrSerLysValAspLysTrpTyr 204
 : : : : :
 2204 ATTTCACAGCAAA.....GGAAATTTATCCAG.....GACTGGATA 2241
 205 GluLeuGlyAsnSerPheLeu.....TyrGlyGlyGlyLagI 218
 : : : : :
 2242 ...ACGGGAAAAACCTGGGATTAAAGATTCTATGTCGTCGATCCAGG 2288
 218 ySerThrLeuThrIleArgLeuArgIleGluThrGlyThrGluProPro 235
 : : : : :
 2289 CGTAGAGTTACCATTCGTTAAAAATCACCC.....AACATCCAGCTG 2332

235 aAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 251
 : : : : :
 2333 TGGCAGTAGTCTGCTGACCTGCTCTTGGAACAAGAGACCTCTGAAGC 2382
 252 Glu.....ProProHisAsnLeuProValProG 261
 : : : : :
 2383 TCCCTCGCTCTCCACCTCTCTCTCCCAAGGAGGAGCCACCCCATC 2432
 261 nLeuThrSerLeuArgProAspLleThrGlnProProSerAsnSerThr 278
 : : : : :
 2433 TCTC.....CCGACTCTTACTCTCACAG 2455
 278 hrGlyLeuIleProThrAsnThrProArg..... 287
 : : : : :
 2456 CCGTGGCGACTGAGCAAACTCCACCGGTGAGAAAAAATTTGTACC 2505
 288 ...AsnSerProGlyValProValLysThrGlyGlnAlaGluPheSerIle 303
 : : : : :
 2506 CTAAACACTCCG.....CCTCCACACAGCGCAGACACTTTTGATCT 2549
 303 uIleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThr 320
 : : : : :
 2550 TGTGACGGGGGCTTCTTAACCTTAATGCTACCAACCGAGGGCCACTG 2599
 320 erSerCysTrpLeuCysLeuSerSerGlyProProTyrTyrGluGlyMet 336
 : : : : :
 2600 AGCTCTTGCGGCTTTGTTGGCCATGGGCCCCCTTATTATGAACATTA 2649
 337 AlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGlnCysTh 353
 : : : : :
 2650 GCCTCATCAGGAGGTGCGCTACTCCACCGAC...CTTGACCGGTGCG 2696
 353 rThrGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyT 370
 : : : : :
 2697 CTGGGGACCAAGGAAGCTCACCTCAGAGTCTTGAGGACAGGCT 2746
 370 hrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTrpSerThr 386
 : : : : :
 2747 TGTGCATAGGAAGGTGCCCTTACCACATCAGACATCTCGCATAGAAC 2796
 387 ValValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAs 403
 : : : : :
 2797 CTATCCATCAATTCCTCCGGAACCATGATGCTGCTCCCTCCACACA 2846
 403 nArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrServ 420
 : : : : :
 2847 TAGCTGTGGGCTTCAGACCTGGCTCACCCCTTGCTCTCCACTGAG 2896
 420 aPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArg 436
 : : : : :
 2897 TTTTATATAGACTAGAGATTTCTGTATCCAGAGTCCAGCTGATTCCTGC 2946
 437 ValTyrTyrHisProGluValValLeuAspGluTyrAspTyrArgTyr 453
 : : : : :
 2947 ATCTATTAATCTATCTGAAGAAAGTTTGTACAGCGCTATGACAAATTC 2996
 453 rAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 470
 : : : : :
 2997 CCCGAGGACTAAAGAGAGGCTGCTCACCTTACCTTACCTGTTTAACTGG 3046
 470 LysLeuGlyThrAlaValAlaGlyValGlyThrGlyThrAlaAlaLeuIleThr 486
 : : : : :
 3047 GGTGGGAATCAGCGCGGGAATAGTACTGTTCACTGCTTAATTTAA 3096
 487 GlyProGlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaAlaMetTh 503
 : : : : :
 3097 GAACCTATAGACCTCCAGCAAGGCTGACAAAGCCCTCAGATCGCATAGA 3146
 503 rGluAspLeuArgAlaLeuLysGluSerValSerLysLeuGluGluSerL 520
 : : : : :
 3147 TGTCACTCTCGGGCTTCACAGACTCAGTCAGAAATTTAAGGAGCTAC 3196
 520 euThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 536

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/046,140
 FILING DATE: 09-MAY-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: FOOT, Brian W.
 REGISTRATION NUMBER: 32,928
 REFERENCE/DOCKET NUMBER: 14538A-003710
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 467-9600
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8655 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-075-272-1

alignment_scores:
 Quality: 1431.00 Length: 727
 Ratio: 3.019 Gaps: 25
 Percent Similarity: 65.199 Percent Identity: 43.466

alignment_block:
 US-09-171-553b-6 x US-09-075-272-1 ..

Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655

```

10 LeuProthArgIgLylGluProLysArgLeuArgIlePro..... 23
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5724 CTTACACGCTCAAGAGGCGCCAGTCAAGATGAAGAAACCCAGCAGAC 5773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
24 ..... LeuSerPheAlaSerIleA 30
5774 AACTGGGCTGTGGAAAGCCAGCATATACCTCTTAAGCTTGCTGCGTGG 5823
30 lAtPheLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeu 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5824 CAGGAGCCCTCCATCACCTCGAC..... 5849
47 lLeaSPSerSerAsnProHisArgProLeuSerLeuThrTrpLeuIleI 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5850 CTAGGAGCACCAACCCCTCATGCTGCCAGTTCAACAGCTCTGGAGAGTCT 5899
63 eaSPProAspThrGlyValThrValAsnSerThrArgGlyValAlaProA 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5900 TAAAT...GAAGAAGGAGACGTTGTATGGGTAGCCACTGCAGTCATCCCC 5946
80 rGgLYThrTrpProGluLeu.....HisPheCys..... 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5947 CTTGGACTGTGGGCTGATCTCACACCCGACATCTGTAAATTACAGCA 5996
91 .....LeuArgLeuIleAsn...Pr 96
5997 GGATCTCCCAATTTGGACCTTCCCGATCATCTGACTTGAATAACCCACC 6046
96 oAlaValLysSerThrProProAsnLeuVal..... 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6047 CTCTGACAAAGTGTGTCCAAACGGGGTAGGAAGCACACTGATGATGTT 6096
107 .....ArgSerTyrGlyPheTyrCysCys 114
6097 CGGGGAGCTTATGAGCTATCTTAGGGCTGCACAGTTTATTTTGGC 6146
115 ProGlyThr.....GluLysGluLysTyrCysGlyGlySerG 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6147 CCTGTGACAGGTGACGAAGAAAGAGCTGCACAAAGATGTAGAGGGGACATC 6196
127 yGluSerPheCysArgArgTrpSerCysValThrSerAsnAspGlyAspT 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

6197 AGACTACTTTTGGCGTAATGACATGTGAACAACAGGGGAGCCCTACT 6246
144 rPLys.....TriProIle..... 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6247 GGAAGCCCTCCGCTGATTTGGAGCTGATTCACGGTAACAACTGGTAGTGT 6296
149 .....SerLeuGlnAspArgValLysPheSerPheValAsnSe 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6297 TATGATAAGCCCAACCAAGAGAAAGAACCCATCAAAATACCTAGATTGC 6346
161 rGLY.....ProGlyLysTyrLysMetMetIle 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6347 TGGGTGTGCTCTTAATAATACAGCCGCCAGGACCA..... 6383
170 yLeuTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLys 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6384 .....TGCMAAGTAATAATACGCAACCC.....CTACTC 6413
187 lIleSerPheThrGluArgLysTyrGlyLysTyrSerLysVal..... 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6414 ATAAATTCACCTGAGAAA.....GGGAAACAAAGCTCCGCTGAGTTGGCT 6457
201 .....AspLysTrpTyrGluLeuGlyAsnSerPheLeuLeuTyrGly 215
6458 TAAAGAAATAGGTGG.....GGTGGCGAGTATATATTCACATTA 6498
215 lYgLYAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6499 GAGACCCCTGGGTATCTTATCTTAGCTAGCTGACAGTAAGA.....GAC 6542
232 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyPr 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6543 CTGGCAGTAACATCATGAGACCCACAAAGAGTCCTTACGGAACAGGCC 6592
248 oProAlaLeuGluProProHisAsnLeuPro...ValProGlnLeuThrS 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6593 CCCAGTTGCACCCGCTCCCGGAGAGTCCAGCGCTCCAGCTCCACCAA 6642
264 eLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyLeu 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6643 CTTACAGGCCC...TACACAGTAGCACCTCATTAGACAAACCTTAGGCC 6689
281 lLeProThrAsnThrProArgAsnSerProGlyValProValLysThrG 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6690 TCCCCACCA.....CTCCTAGATACAGA 6712
297 yGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSer 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6713 AAACCGTCTGCTCAGTCTAGTTCAGGAGCCCTTTTGTATTAAATAGCA 6762
314 hrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyPro 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6763 CTAATCTTAATATGATCAATCATGTTAGTATGCTTATGCTCTTAACCCC 6812
331 ProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysG 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6813 CCTATTATGAAAGAAATGCTCAGACATGAACATTAATTAATCTTCAGA 6862
347 uHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThr 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6863 TCAT...TCTCAATGTCTTTGGGAGAGAACAGAGAACTTACCTGACAG 6909
364 lValSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisGln 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6910 CAGTTTACGAAATGGGCTTGTATTAGTAGGCGCCAGCAATAATG 6959
381 HisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnT 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6960 CACCTCTGTAAACGAGCTCAAAATATCCGACCTAACAAAGGTGTGAGTA 7009
397 rLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThr 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7010 TCTAGTCCCTCCCATAGACACAGTATGGGCTTGCAATACAGTCTACCTC 7059

```

```

414 roCysValSerThrSerValPheasnGlnSerLysAspPheCysValMet 430
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7060 CTTGATATCTATCTGTTTTCATAGCTCCAAAGATTCTGTATTTA 7109
431 ValGlnIleValProArgValTyrTyrHisProGlnGluValValLeuAs 447
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7110 GTTCAGCTTATCTCTAGACCTCCCTATCATGATGATAGTACGATTTCTAGA 7159
447 pGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeuT 464
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7160 CAAATTTGAACATCGGCTC...CCCTGGAAAAGAGAACCCATTACTTTAA 7206
464 hrIleuAlaValMetLeuGlyLeu...ThrAlaValGlyValGlyThr 479
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7207 CGTTAGCAGCTGTTTAGAGATGGAGCTGCCGAGCTGGGGTAGGTA 7256
480 GlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGlnLysGlyLeuG 496
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7257 GGAACCGCTGCTTATCCAGACCCCGCATCTTTCAG... 7295
496 yGluLeuHisAlaIleMetThrGluAspLeuArgAlaLeuLysGluSerV 513
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7296 .GAATTACGTACACTATGATCTCAGAGCTATAGAACACACTCTA 7344
513 alSerAsnLeuGlnGluSerLeuThrSerLeuSerGluValValLeuGln 529
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7345 TACCAAACTAGAGAAATCTTTAACTTCTCTGCTGAGGTAAGTACTCAA 7394
530 AsnArgArgGlyLeuAspLeuLeuPheLeuArgGlnGlyLeuCysAl 546
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7395 AATAGAGAGAGATTAGATTGTATTCTTAAGAAGAGAGACTTTGTGC 7444
546 alAlaLeuLysGlnGluCysPheTyrValAspHisSerGlyAlaIleA 563
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7445 TGCCTCAAGAGAGAAATGTTCTTCTATGTCGACCACTCGGAGATGATCA 7494
563 rGaSPSerMetAsnLysLeuArgLysLysLeuGluArgArgArgGlu 579
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7495 AAGATTCTATGCCCAACTTAGAAGACGCTGATATACGTCACAAAGAGAG 7544
580 ArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgSerProT 596
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7545 AGAGAGAGACAAACAGATGTTGAAGCTGTTAATAGTCCCTGC 7594
596 pMetThrIleuLeuSerAlaLeuThrGlyProLeuValValLeuLeuL 613
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7595 GCTCACCACTCTCTTCCACTATAGCGGGGACCTTGATTAATCTCTGC 7644
613 euleuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheVal 629
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7645 TTTTGCTTACTTTGGCCCTGCATCTCTATATAGTAGTAGCCCTTAT 7694
630 ArgGluArgValSerAlaValGlnIleValValLeuArgGlnGlnTyrG 646
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7695 AGAGAAAGAAATAAATGCAGTACAGTTATGCTATTAACAACAATATCA 7744
646 nGlyLeuLeuSerGlnGlyGluThrAspLeu 656
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
7745 G...GTCTTCCAGAGGCCGAAACCTGCTC 7772

```

seq_name: /cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-110-300A-8

```

; seq documentation block:
; Sequence 8, Application US/08110300A
; Patent No. 5643756
; GENERAL INFORMATION:
; APPLICANT: Pinter, Abraham
; APPLICANT: Kayman, Samuel
; TITLE OF INVENTION: FUSION GLYCOPROTEINS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Davis Hoxie Faithfull and Hapgood

```

```

; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,300A
; FILING DATE: 20-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jacobs, Seth H.
; REGISTRATION NUMBER: 32,140
; REFERENCE/DOCKET NUMBER: 11698A50
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-757-2200
; TELEFAX: 212-586-1461
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-110-300A-8

```

alignment_scores:

Quality	Length
1299.00	709
Ratio: 2.855	Gaps: 26
Percent Similarity: 64.175	Percent Identity: 42.595

alignment_block:

US-09-171-553b-6 x US-08-110-300A-8

Align seg 1/1 to: US-08-110-300A-8 from: 1 to: 8323

```

4 ThrLeuSerArgArgHisLeuProThrArgGlyGluProLysArgLe 20
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5787 ACCCTCTCAAAATCC...CCTAAAGATAGATTGACCGGGGAGCT 5830
20 uArgIleProLeuSerPheAlaSerIleAlaIleThrPheLeuThrI 37
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5831 CCTAATCCCTTAATCTC...TTCTGTCTCTC... 5861
37 leThrProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHis 53
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5862 .....AAAGGGCCAGATCCGACGACCCGCTCCAGCCCTCAC 5900
54 ArgProLeuSerLeuThrTyrPheLeuIleIleAspProAspThrGlyValTh 70
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5901 CAGGTCTACACATTACTGCGAAGTACACCAATGGGATCGG...GAGAC 5947
70 rValAsnSerThrArgGlyValAlaProArgGlyThrPheProGluL 87
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5948 AGTATGGGCAATATCAGGCAACCCCTCTGTGACTGTGTGCTGCTC 5997
87 eu.....Hispe... 89
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
5998 TCACCCACAGATTGTGTATGTAGCTCTCAGTGGGCCGCCCACTGGGG 6047
89 ..... 89
6048 CTAGAGTATAGGCCCTATTCTCGCCCGGGGCGCCCTGTGTGCTC 6097
90 .....CysLeuArgLeuIleAsnProAlaVal 98
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

6098 AGGAGCAGCGGGAGAGCTTGACAGCGTGTCCAGAGACTGCACAGCGACCT 6147
98 allysSerThrProPro.....AsnLeuValArg... 107
6148 TGACCTCTCCACACCCCTCGGTGCACACATGCCGTGAGACAGACTTAAGTGC 6197
108SerrTyrGlyPheTyrCysCysProGlyTTh 117
6198 GACACAGTAACTCATTAATCACTAGTAGGGATTTTATGTCTGCCCGCGGTCT 6247
117 rGluLys.....GluLysTyrCysGlyLysSerGlyLysSerPheC 131
6248 ACATCGCCCGCCGGAGACCAAGCTCGTGGGGTGCACAGCTCCTCTACT 6297
131 ysarGhrGtTPSerCysValThrSerAsnAspGlyAspTTrpLysTTrpPro 147
6298 GTCCCTCTTGGGGCTCGAGAACCCGAGTAGAGTACTGAGG...CCG 6344
148 lIleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProGly 164
6345 TCCTCTTTTGGACACTACATCA.....GTAGCAACAATCTCAC 6385
164 yLysTyrLysMetLeuLysLeuTyrLysAsp...LysSerCysSerProS 180
6386 CTCTAACCAACGAGCTGTACAGTATGCAAGACATTAAGTGGTGCATGCC. 6434
180 eArspLeuAspTyrLeuLyslIleSerPheThr..GluArgLysThrGlyL 196
6435TTGGCTATCCGGTTTACAAACGCCGGGAAACAGCTC 6470
196 ysTyrSerLysValAspLysTTrpTyrGluLeuGlyAsnSerPheLeu..Le 212
6471 ACCTC.....ATGGCAACACTGACACATATTTGGGGATCTACACTTTTATGT 6514
212 uTyrGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgLysLeuThr 229
6515 CTCTGGACAGACGCCAGGCTTACTTTGGGATCCGACTCAGTATATCAA. 6563
229 hGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
6564 ..AMCTAGAGACTGAGATCCCAATAGACCAAAACCCGCTGTCGGCAGAC 6611
246 GlnGlyProProAlaLeuGlnProProHisAsnLeuProValProGlnLe 262
6612 CAA.....CTTTGCTTCCCGCTACCTAATCCCTACCC..... 6644
262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrc 279
6645AAACGTGCAGAGTGTCCCGCGCTCTAGTTGCAGTCCCA 6684
279 lYleuIle...ProThrAsnThrProArgAsnSerProGlyValProVal 294
6685 CATGTATTTTCCCGCTGCCCATCCACTCCAGTCCGCCG.....CCAGCA 6728
295 LysThrGlyGlnArgLeuPheSerLeuLeuGlnGlyAlaPheGlnAlaI 311
6729 GGAAGCGGAGAGATTAATCTAAATCTACTACAGGAGACTTACACAGGACT 6778
311 eaSnSerThrAspProAspAlaThrSerSerCysTTrpLeuCysLeuSerS 328
6779 CAACCTTACCAACCTGTATAAATCACTCAAGAAGTGTGCTGTTATGCTTACTGT 6828
328 eArgLysProProTyrTyrGlnGlyMetAlaLysGluArgLysPheAsnVal 344
6829 CTGGACCCCGCATTAACAGAGGGGTGGCCGCTAGGTACTTATTCAC 6878
345 ThrLysGlnHisArgAsnGlnCysThrTTrpGlySerArgAsnLysLeuTh 361
6879 CATACCTCTGCCCAAGCTAACGTGCTCGTGGCCCTCCCAACACAGCTGAC 6928
361 rLeuThrGlnValSerGlyLysGlyThrCyslIleGlyLysAlaProProS 378
6929 CCTGTCCCAAGTGAAGTACGCGGAGCTCTGTGATGAGAACAGTCCCAAAA 6978

```

395 AsnGluTyrLeuValProGlyTyrAsnArgTyrPrtAlaCysAsnThrG1 411
    :: ||||| :: |||||
7023 TCTACTATCTAGTTGGCCCCACAGACATAGTATGGGCGATGAACACTGG 707
    ||||| |||||
411 yLeuThrProGlyValSerThrSerValPheAsnGlnSerIysAspPheC 428
    ||||| |||||
7073 ACTACTCCAGGCCCTATCTGGCACCGCTCTTAATGACACCACTAGTATT 712
    ||||| |||||
428 yValMetValGlnIleValProArgValTyrTrpHisProGluVal 444
    ||||| |||||
7123 GCGTTCGCTGGAAATATAGGCCACAGGCTACCTACCATCTCCCACTTAC 717
    ||||| |||||
445 ValLeuAspGluTyrAspTyrArgTyrTrpAsnArgProIysArgGluProVa 461
    ||||| |||||
7173 GTCTATAGCCAGCTTTGAA...AAATCCATGACATTAAGAAGAACAGT 721
    ||||| |||||
461 lSerLeuThrLeuAlaValMetLeuGly.....LeuGlyThrA 474
    ||||| |||||
7220 GTCCCTTAACCTTGGCCCTTATTATTAGGTGGCTTAACATATGGGTGCATCG 726
    ||||| |||||
474 lValGlyValGlyThrGlyThrAlaIleLeuIleThrGlyProGlnGln 490
    ||||| |||||
7270 CCGGGGAGATAGGACACAGAACATACCGCCCTGGTC..GCCACCCAGCAG 731
    ||||| |||||
491 LeuGluLysGlyLeuGlyGluLeuHisAlaIleMetThrGluAspLeuArg 507
    ::::: :::::
7317 TTTCAG.....CAGCTCCATGCTCGCTGACAAAGATGATGATCA 735
    ||||| |||||
507 gAlaLeuLysGlnSerValSerAsnLeuGluGlnSerLeuThrSerLeuS 524
    ::::: :::::
7355 AGAAGTGGAAAGTCAATTACTTAACCTAAGAAAGTCTCTTCACTTGTGT 740
    ||||| |||||
524 erGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArg 540
    ||||| |||||
7405 CTGAGGTTGACTGCAACAATGACGACAGGCCCTAGACCTGTTCTCTTA 745
    ||||| |||||
541 GluGlyGlyLeuCysAlaIleLeuLysGluGluCysCysPheTyrValAs 557
    ||||| |||||
7455 GAGGAGAGAGACTGTGCTGCCCTAAAGAAAGATGTTTCTTATGCTGA 750
    ||||| |||||
557 pHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 574
    ||||| |||||
7505 CCAATACAGGCTTGTAAAGATATGATAGCCCAAAATTAAGAGAGACCTT 755
    ||||| |||||
574 lIuArgArgArgArgGluArgGluAlaIleAspGlnGlyTrpPheGluGly 590
    ::::: |||||
7555 CTGAGAGACAAAACATTTTGGAGTCGAGCCAGAGATGTTCTGAAGATGG 760
    ||||| |||||
591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyPr 607
    ||||| |||||
7605 TTTAACAGATCCCCGCTGTTTACACGCTGATATCCACATCATGAGGGCC 765
    ||||| |||||
607 OleuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIleAsnA 624
    ||||| |||||
7655 TCTCATTTATCTCTACTAATTTGCTTTTGGACCTCGATTCCTTATTC 770
    ||||| |||||
624 rrpPheAlaIlePheValArgGluArgValSerAlaValGlnIleMetVal 640
    ||||| |||||
7705 GATTAGTTCATTTGTTTAAAGACAGATCTCAGTAGTCCAGGCTTTAGTC 775
    ||||| |||||
641 LeuArgGlnGlnTyrGlnGlyLeu 648
    ||||| |||||
7755 CTGACTCAACAATATACACACAGTA 7778
    ||||| |||||
seq_name: /cgn2_6/p/odata/2/1na/5B_COMB.seq:US-08-886-642-8
: Sequence 8, Application US/08886642

```

```

Patent No. 5952474
GENERAL INFORMATION:
APPLICANT: Pinter, Abraham
APPLICANT: Kayman, Samuel
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 45 Rockefeller Plaza, Suite 2800
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/886,642
FILING DATE: 01-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/110,300
FILING DATE: 20-AUG-1993
APPLICATION NUMBER: 07/938,100
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hone, William J.
REGISTRATION NUMBER: 26,739
REFERENCE/DOCKET NUMBER: 0763/010002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212/765-5070
TELEFAX: 212/258-2291
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-886-642-8

alignment_scores:
Quality: 1299.00 Length: 709
Ratio: 2.855 Gaps: 26
Percent Similarity: 64.175 Percent Identity: 42.595

alignment_block:
US-09-171-553b-6 x US-08-886-642-8 ..
Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

4 ThrLeuSerArgArgHisLeuProThrArgGlyGluProLysArgLe 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5787 ACGCTCTCAAAATCC.....CCTAAAGATAGATTGACCCGGGAGACT 5830
20 uArgTLeuProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeu 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5831 CCTAATCCCTTAATCTC.....TTCTGTCGTC.... 5861
37 LeuThrProGlnAlaSerLysArgLeuIleAspSerSerAsnProHis 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5862 .....AAAGGGCCAGATCCGCGACGCCGCGCTCCAGCCCTCAC 5900
54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspTrpGlyVal 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5901 CAGCTCTAACAATCTACTGCGAAGTGCACCAATGGGATCG...GAGAC 5947
70 ValAsnSerThrArgGlyValAlaIleProArgGlyThrTrpTrpProG 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5948 AGTATGGCAATATTCAGGCAACCAACCCCTGTGGACTGTGGCCAGTCC 5997

```

```

87 eu.....Hispe... 89
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
5998 TCACCCAGATTGTGTATGTAGCTCTCAGTGGCGCCGCCCACTGGGG 6047
89 ..... 89
6048 CTAGAGTACAGGCCCTTATCTCGCCCCGGGGCCCCCTTGTGCTC 6097
90 .....CysLeuArgLeuIleAsnProAlaVal 98
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6098 AGGAGCAGCGGAGACCTTGACAGCTGTGCCAGAGACTGCACAGCCCT 6147
98 aLysSerThrProPro.....AsnLeuValArg... 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6148 TGACCTTCCCTCACCCTCGGTGCACACACTGCTGAAACAGACTTAAGCTG 6197
108 .....SerTrpGlyPheTrpCysCysProGln 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6198 GACCAGGTACTATTAATCAAGTGAAGGATTTATGTGTGCCCCGGGTC 6247
117 rGluLys.....GluLysTrpCysGlyGlySerGlyLysSerPheC 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6248 ACATCGCCCGCCGGAAGCAAGTCTGTGGGGGTCCAGACTCTTCTACT 6297
131 ysArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpPro 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6298 GTGCCCTTGTGGGCTGCGAGACAACCGGTAGAGTACTGTAGAG...CCG 6344
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyProG 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6345 TCCTCTTCTTGGGACTCATACACA.....GTACACAACATCTCAC 6385
164 yLysTrpLysValAspLysTrpTrpGluLeuGlyAsnSerPheLeu 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6386 CTCTAACCAAGGCTTTCAGGTATCCAAAGCAATAGTGTGCAATGCC. 6434
180 eAspLeuAspTrpLeuLysIleSerPheThr..GluArgLysThrGly 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6435 .....TTGGCTATCCGTTTACAAACCGCGGGAACACAGTCC 6470
196 ySTySerLysValAspLysTrpTrpGluLeuGlyAsnSerPheLeu 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6471 ACCTC.....ATGGACAACCTGGACACTATTGGGGCTTCTTATATGT 6514
212 uTrpGlyGlyGlyAlaGlySerPheLeuThrIleArgLeuArgIleGlu 229
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6515 CTCGTGACAGAGCAAGCGCTTACTTTCGGATCCGATCAGTATACAA. 6563
229 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6564 ..AATCTAGACCTCGATCCCAATAGACCAACCCGCTCTCGACGAC 6611
246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGlnLe 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6612 CAA.....CTTCTGTTCCCGTACTTATCCCTACCC..... 6644
262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrG 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6645 .....AAACCTGCACAGTCTCCCCCGCTCAGTTCGACTCCCA 6684
279 ILeuLeuIle..ProThrAsnThrProArgAsnSerProGlyValProVal 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6685 CATGATTTCCCGTCCCGCCACTCCACTCCACTCGACCCCG.....CCAGCA 6728
295 LysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaI 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6729 GGAAGGGAGACAGATTACTTAATCTAGTACAGGAGCTTACCGAGCACT 6778
311 eAsnSerThrAspProAspAlaThrSerSerCysTrpLeuGlyLeuSer 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6779 CAACCTTACCAACCCCTGTAATACTCAAGAGTGTGCTTAACTGCTAGTGT 6828
328 eTrpGlyProProTrpTrpGlyGluMetAlaLysGluArgLysPheAsnVal 344

```



```

6829 CTGGACCCCTATTAGAGGGGGTCCCTCTAGTACTATTATTCAC 6878
345 ThLysGluHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeu 361
6879 CATACCTCTGCCACGTAACCTCTCGTGGCCCTCCACACAGAGTAC 6928
361 rLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAlaPro 378
6929 CCTGTCCGAGTGTAGTACGAGGGGCTGTCATAGGAGAACGCCAAA 6978
378 eHisGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSer 394
6979 CTCACAGAGCCCTGTCACACACTACCTT...AAGCAGAGCAAGGG 7022
395 AsnGlyTyrLeuValProGlyTyrAsnArgTrpTyrPalaCysAsn 411
7023 TCTTACTATCTAGTTGCCCCACAGAGAACTATGTGGCATGTAAC 7072
411 yLeuThrProCysValSerThrSerValPheAsnGlnSerLysAsp 428
7073 ACTCCTCATGCTATCTGTCCACCGCTGTAATCGCACCACTGACT 7122
428 yValMetValGlnIleValProArgValTyrTyrHisProGluGlu 444
7123 GGTCTCTGTAATATGAGGAGGTACCTACCATCTCCCGATTAC 7172
445 ValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGlu 461
7173 GTCTATAGCCAGTTGAA...AAATCCATATGACATAAAGAGAC 7219
461 lSerLeuThrLeuAlaValMetLeuGly...LeuGlyThr 474
7220 GTCCCTTAACCTGGCCTTATTATTAGGGGCTAAGTGGGTCATG 7269
474 lValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 490
7270 CCGCGGAGTAGGAGAGAACTACCGCGCTGTC...GCCACCCAGAG 7316
491 LeuGlyLysGlyLeuGlyGluHisAlaIleMetThrGluAspLeu 507
7317 TTTCAG...CAGCTCCATGCTCGCGTACAGATGATCTCA 7354
507 gAlaLeuLysGlnSerValSerAsnLeuGluSerLeuThrSerLeu 524
7355 AGAAGTCAAAAGTCAATTACTACCTGAAAGTCTTCTCTGTTGT 7404
524 eGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeuArg 540
7405 CTGAGGTGTACTGCGAATCGAGAGGCTAGACCTGTCTCTAAA 7454
541 GluGlyGlyLeuGlyAlaIleLeuLysGluLysCysPheTyrVal 557
7455 GAGGAGAGACTGTGTGCTGCCCTAAAGAAAGATGTTCTATGCTGA 7504
557 pHisSerGlyAlaIleArgAspSerMetLysLysLeuArgLysLeu 574
7505 CCATACAGCGCTAGTAGAGTACTATGCGCAATTAAGAGAGACTCT 7554
574 lValArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGly 590
7555 CTCAGAGCAAAACATTTTGATGAGCCAAAGATGTTGGAAGATG 7604
591 PheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThr 607
7605 TTTTACAGATCCCTGCTTTTACACAGTGTATATCCACATCATG 7654
607 OLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeuIle 624
7655 TCTCATATATCTCTACTAATTCGCTTTTGGACCTGCATCTTAT 7704
624 rGpIleValAlaPheValArgGluArgValSerAlaValGlnIle 640

```

```

7705 GATTAGTCAATTGTGTAAGACAGATCTGATGTCAGGCTTACTC 7754
641 LeuArgGlnIleTyrGlnGlyLeu 648
7755 CTGACTCAACATATACACAGCTA 7778

seq_name: /cgn2_6/ptodata/2/lna/PCTus_COMB.seq:PCT-US93-08041-8
seq_documentation_block:
  ; Sequence 8, Application PCT/US9308041
  ; GENERAL INFORMATION:
  ; APPLICANT: The Public Health Research Institute of the City of
  ; APPLICANT: New York, Inc.
  ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
  ; NUMBER OF SEQUENCES: 16
  ; CORRESPONDENCE ADDRESS:
  ; ADDRESSEE: Davis Hoxie Faithful and Hapgood
  ; STREET: 45 Rockefeller Pl.
  ; CITY: New York
  ; STATE: New York
  ; COUNTRY: USA
  ; ZIP: 10111
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE: Floppy disk
  ; COMPUTER: IBM PC compatible
  ; OPERATING SYSTEM: PC-DOS/MS-DOS
  ; SOFTWARE: PatentIn Release #1.0, Version #1.25
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER: PCT/US93/08041
  ; FILING DATE:
  ; CLASSIFICATION:
  ; PRIORITY APPLICATION DATA:
  ; APPLICATION NUMBER: US 07/938,100
  ; FILING DATE: 28-AUG-1992
  ; PRIORITY APPLICATION DATA:
  ; APPLICATION NUMBER: US 08/
  ; FILING DATE: 20-AUG-1993
  ; ATTORNEY/AGENT INFORMATION:
  ; NAME: Jacobs, Seth H.
  ; REGISTRATION NUMBER: 32,140
  ; REFERENCE/DOCKET NUMBER: 11698450
  ; TELECOMMUNICATION INFORMATION:
  ; TELEPHONE: 212-757-2200
  ; TELEFAX: 212-586-1461
  ; INFORMATION FOR SEQ ID NO: 8:
  ; SEQUENCE CHARACTERISTICS:
  ; LENGTH: 8323 base pairs
  ; TYPE: nucleic acid
  ; STRANDEDNESS: double
  ; TOPOLOGY: linear
  ; MOLECULE TYPE: DNA (genomic)
  ; HYPOTHEICAL: NO
  ; ANTI-SENSE: NO
  ; PCT-US93-08041-8

alignment_scores:
  Quality: 1299.00      Length: 709
  Ratio: 2.855          Gaps: 26
  Percent Similarity: 64.175  Percent Identity: 42.535

alignment_block:
  US-09-171-553b-6 x PCT-US93-08041-8 ..
  Align seg 1/1 to: PCT-US93-08041-8 from: 1 to: 8323

4 ThLeuSerArgArgHisLeuProThrArgGlyGluProLysArgLe 20
5787 ACAGCTCTCAAAATCC...CTTAAGATAGATTGACCGCGGAGCT 5830
20 uArgIleProLeuSerPheAlaSerIleAlaIlePheLeuThrIle 37
5831 CCTATCCCTTAATTC...TTCCGTCTCTC... 5861

```

37 LeThrProGlnIAserSerIysTrpLeuIleAspSerSerAsnProHis 53
5862 AAGGGGCGAGATCCGACGACGCCCGGCTCCAGCCTCAC 5900
54 ArgProLeuSerLeuThrTrpLeuIleIleAspProAspThrGlyVal⁷⁰
5901 CAGGCTACAACTTACTCGGGAAGTGCACCAATGGGATCGG...GAGAC 5947
70 rValAsnSerThrArgGlyValAlaProArgGlyThrTrpTrpGlnL 87
5948 AGTATGGGCAATATCAGGCAACCAACCCCTGTGGACTTGGTGGCCAGTCC 5997
87 eu.....Hispe... 89
5998 TCACCCGAGATTGTGTATGTTAGCTCTCAGTGGCGCCGCCACTGGGG 6047
89 89
6048 CTAGAGTATCAGGCCCTATTCTCGGCCCGGGGCCCTTGTTGCTC 6097
90CysLeuArgLeuIleAsnProAla⁹⁸
6098 AGGAGACACGGGAACGTGCGAGCTGTGCCAGAGACTGCACACAGCCCT 6147
98 allYserThrPro.....AsnLeuValArg... 107
6148 TGACCTCCTCACCCCTGGTGCAACACTGCCCTGGAACAGACTTAACCTG 6197
108SerTrpGlyPheTrpCysCysProGly¹¹⁷
6198 GACCAAGTAACTCAATAATCAAGTGGGATTATGTCGCCCGGCTC 6247
117 rGluLys.....GluLysTrpCysGlyGlySerGlyGluSerPhe¹³¹
6248 ACATCGCCCCCGGGAAGCCAGTCTGTGGGGTGCAGACTCCTTCACT 6297
131 ysaTrpArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrp¹⁴⁷
6298 GTGCTCTTGGGGCTGCGAGACACCGGTAGAGTATACTGGAC...CCC 6344
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSerGlyPro¹⁶⁴
6345 TCCTCTCTGGGACTACATCA.....GTAGCAACAATCTCAC 6385
164 yLysTrpLysMetMetLysLeuTrpLysAsp...LysSerCysSerPro¹⁸⁰
6386 CTCATACCAAGCTGTACAGTATGCAAAGACAATAGGTGTCATGCC. 6434
180 eArspLeuAspTrpLeuLysIleSerPheThr...GluArgLysThr¹⁹⁶
6435TTGGCTATCCGGTTTCAAAACGCGGGAACAGCTC 6470
196 ystYrSerLysValAspLysTrpTrpGlyLeuGlyAsnSerPheLeu²¹²
6471 ACCTC.....ATGGACAACCTGACACTATTGGGGTCTACCTTTATGT 6514
212 uTrpGlyGlyGlyAlaGlySerThrLeuThrIleArgLeuArgIle²²⁹
6515 CTCGTGACAGGACCCAGGCGCTTACTTTCGCGATCCGACTCAGTTATCAA. 6563
229 hrcLysThrGluProProValAlaMetGlyProAspLysValLeuAla²⁴⁵
6564 ...AATCTAGGACCTGGATCCCAATAGACCAACCCCGTCTCTGGACAC 6611
246 GlnGlyProProAlaLeuGluProProHisAsnLeuProValPro²⁶²
6612 CA.....CTTTCGTTCCCGCTACCTAATCCCTTACC 6644
262 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThr²⁷⁹
6645AAACGTGCCAAGTCTCCCGCGCTCTAGTTGCAGCTCCA 6684


```

159 aIAsnSerGlyProGlyLysTyrLysMetLysLeuTyrLysasp... 174
1944 TAAACAACAATCTCACCTGTGACGAGCTGTGACAGATGATGCAAAAGATAT 1993
175 LysSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrG1 191
1994 AAGGTGGTCAACCCC.....TTAGTTATTGGGTTACAGAA 2028
191 u.....ArgLysThrGlyLysTyrSerLysValAspLysTyrGluL 206
2029 CGCGGGAGACGGGTACTCTCGGACACAGACATTACTG..... 2071
206 euGlyAsnSerPheLeuLeuTyrGlyGlyAlaGlySerThrLeuThr 222
2072 ..GGCTTACGTTTGTATGTCCTCGACAAAGATCCAGGGCTTACATTGGG 2119
223 IleArgLeuAspTyrLeuThrGlyThrGluProProValAlaMetGlyPr 239
2120 ATCCGACTCAGATACCA...AATCTAGAGACCCGCGTCCCAATVAGGCC 2166
239 oAspLysValLeuAlaGluGlnGlyProProAlaLeuGluProProHisA 256
2167 AAACCCCGTTCGGACAGACCAAGCCA.....CTCTCCAAGCCCAAA. 2209
256 snLeuProValProGlnLeuThrSerLeuArgProAspIleThrGlnPro 272
2210 ....CCTGTTAAGTCG.....CCTTCAGTCACCAACCA 2239
273 ProSerAsnSerThrThrGlyLeuIleProThrProAlaGlnAsnSe 289
2240 CCAGAGGG.....ACTCCTCTCTCCCTACCCCACTTCCA..... 2275
289 rProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGln 306
2276 .....CGCGGGAGACGAAATVAGGTGCTAATCTACTAGACG 2315
306 lValAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCys 322
2316 GAGCCTACCAAGCCTCAACCTCAGACAGTCTGACAAACCAAGAGTGC 2365
323 TrpLeuCysIleSerSerGlyProProTyrTyrGlnGluMetAlaLysG1 339
2366 TGGTGTGTCTAGTAGCGGAGCCCTACTACTACGAAGGGTTCGCTCC 2415
339 uArgLysPheAsnValThrLysGlnHisArgAsnGlnCysThrTrpLys 356
2416 GGGTACCTACTCCACCATACCTCTGCTCAGCCACAGTGCCTGGCCT 2465
356 eArgAsnLysLeuThrLeuThrGluValLysGlyLysGlyThrCysIle 372
2466 CCCAACACAAAGTTGACCTGTCCGAAAGTACCGGAGGAGGAGCTTCGATA 2515
373 GlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTy 389
2516 GGAGCAGTTCCCAAAACACATCAGCCCTATGATATACCAACCAAGACAAG 2565
389 rGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrp 406
2566 CAGTGGAGGGTCC.....TATATCTAGTTGCCCTTACAGGTACCATGT 2609
406 rPalAcysAsnThrGlyLeuThrProCysValSerThrSerValPheAsn 422
2610 GGGGTGTAGTACCGGGCTTACTCATCTCCATCTCCACCATCTGTAAC 2659
423 GlnSerLysAspPheCysValMetValGlnIleValProArgValTyrTy 439
2660 CTTCACACTGATATGTTGTTCTGTGCAACTCTGGCCAGAGTCACTA 2709
439 rHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgp 456
2710 TCATTTCCCCCAAGCTATGTTTACGGCTGTTTGAG...AGATCCAAACCGAC 2756
456 rOlyAspGluProValSerLeuThrLeuAlaValMetLeuGly..... 470
2757 ACAAAAGACACCGGTGCTTACCTGACCTGACCTATATATGGGTGACTA 2806
471 .....LeuGlyThrAlaValGlyValGlyThrGlyThrAlaLeu1 485
2807 ACCATGGGGGAATTCGCGTGAATVAGCAACAGGAGTACTAGCTCTAAT 2856
485 eThrGlyProGlnGlnLeuGlnLysGlyLeuGlyGlnLeuHisAlaIam 502
2857 G...GCCACTCAGCAATTCAG.....CAGCTCCAAAGCCGAG 2891
502 eThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluL 518
2892 TACAGGATATCTCAGAGAGGTTGAAAATCAATCTTACTTAAAG 2941
519 SerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAs 535
2942 TCTCTCACTCCCTGCTGAAGTTGCTCCTACAGAAATCGAAGGGGCTTGA 2991
535 pleuLeuPheLeuArgGluGlyLeuCysAlaAlaLeuLysGluGluC 552
2992 CTGTATTCTTAAAGAAAGAGGCTGTGTCTCTTAAAGAAAGAAAT 3041
552 yScysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLys 568
3042 GTTGCTTATGCGGACGACACAGACACTAGTACAGACAGATGCCCAA 3091
569 LeuArgLysLysLeuGluArgArgArgGluArgGluAlaAspGlnG1 585
3092 TTGAGAGAGAGGCTTAATCAGACAGAACTGTTTGAGTCACTCAAG 3141
585 yTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeu 602
3142 ATGGTTTGAGGACGCTTTACAGATCCCTTGGTTTACCACTTGATAT 3191
602 eAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGly 618
3192 CTACCATATATGGACCCCTCATTTGATCTCATATGATTTCTCTTGGA 3241
619 ProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAl 635
3242 CCTGCATTCCTTAATVCAATVGTTCATTTGTAAAGACAGATCTCAGT 3291
635 aValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 648
3292 AGTCCAGGCTTAGTCTGACTCAACATATCCACAGCTA 3331

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 02:58:34 ; Search time 212.72 Seconds

(without alignments)
3147.184 Million cell updates/sec

Title: US-09-171-553B-9

Perfect score: 2956

Sequence: 1 tgcctttaggttaggaac.....aaaaaaaaaaaaaaaaaaaaa 2956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfillseq.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2889.4	97.7	3482	3	US-09-111-085-3
2	1882.8	63.7	7333	4	US-08-766-528-2
3	1563.4	52.9	2462	3	US-09-111-085-1
4	1489.2	50.4	8132	4	US-08-766-528-3
5	1231.8	41.7	8060	4	US-08-766-528-1
6	973.8	32.9	2000	4	US-09-376-781-25
7	973.8	32.9	2000	4	US-09-376-781-30
8	710.4	24.0	1493	4	US-09-376-781-24
9	455	15.4	6312	4	US-09-011-745-7
10	451.2	15.3	8535	3	US-08-716-351A-1
11	372.2	12.6	8655	3	US-09-075-272-1
12	357.8	12.1	10970	3	US-08-716-351A-5
13	343	11.6	638	4	US-09-376-781-18
14	307	10.4	6028	4	US-09-011-745-5
15	303.8	10.3	1965	1	US-08-258-420-9
16	301.4	10.2	2499	1	US-08-105-483-310
17	301.4	10.2	2499	1	US-08-709-209-310
18	301.4	10.2	2499	1	US-08-458-101-310
19	300.6	10.2	2001	3	US-08-850-961-13
20	298.8	10.1	8323	1	US-08-110-300A-8
21	298.8	10.1	8323	2	US-08-886-642-8
22	298.8	10.1	8323	5	PCR-US93-08041-8
23	298.2	10.1	1911	1	US-08-258-420-8
24	298.2	10.1	1911	1	US-08-258-420-13
25	296.8	10.0	8202	4	US-08-552-369-19
26	293.4	9.9	3925	4	US-09-011-745-9
27	290.2	9.8	1914	1	US-08-258-420-7

ALIGNMENTS

28	290	9.8	6061	4	US-09-011-745-6	Sequence 6, Appli
29	283.6	9.6	8332	3	US-08-850-961-1	Sequence 1, Appli
30	269.6	9.1	276	4	US-09-376-781-8	Sequence 8, Appli
31	253.4	8.6	278	4	US-09-376-781-10	Sequence 10, Appli
32	192	6.5	5109	1	US-08-073-836-1	Sequence 1, Appli
33	192	6.5	5176	1	US-08-073-836-2	Sequence 2, Appli
34	157.2	5.3	10306	3	US-08-716-351A-4	Sequence 4, Appli
35	144.4	4.9	9661	4	US-09-376-781-7	Sequence 3, Appli
36	128.2	4.3	445	4	US-08-484-126-4	Sequence 4, Appli
37	113	3.8	10367	2	US-08-110-300A-9	Sequence 7, Appli
38	113	3.8	10367	2	US-08-886-642-9	Sequence 9, Appli
39	113	3.8	10367	5	PCR-US93-08041-9	Sequence 9, Appli
40	108.2	3.7	1237	1	US-08-798-000-2	Sequence 9, Appli
41	105.8	3.6	1374	1	US-08-278-630A-9	Sequence 2, Appli
42	103.8	3.5	1356	2	US-08-484-126-4	Sequence 9, Appli
43	102.8	3.5	735	1	US-08-798-000-3	Sequence 4, Appli
44	102.8	3.5	1234	1	US-08-798-000-1	Sequence 3, Appli
45	102.8	3.5	7308	4	US-09-011-745-3	Sequence 3, Appli

RESULT 1	US-09-111-085-3	Sequence 3, Application US/09111085
Patent No. 6100034		
GENERAL INFORMATION:		
APPLICANT: Stoye, Jonathan P		
APPLICANT: Weiss, Robin A		
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope		
TITLE OF INVENTION: Specific sequences		
FILE REFERENCE: 4238/75168		
CURRENT APPLICATION NUMBER: US/09/111.085		
CURRENT FILING DATE: 1998-07-07		
EARLIER APPLICATION NUMBER: GB 9710154.7		
EARLIER FILING DATE: 1997-05-16		
NUMBER OF SEQ ID NOS: 16		
SOFTWARE: PatentIn Ver. 2.0		
SEQ ID NO 3		
LENGTH: 3482		
TYPE: DNA		
ORGANISM: Porcine retrovirus		
US-09-111-085-3		
Query Match	97.7%	Score 2889.4; DB 3; Length 3482;
Best Local Similarity	99.3%	Pred. No. 0;
Matches 2913; Conservative	0;	Mismatches 21; Indels 1; Gaps 1;
1 tgcctttaggttaggaacacccctggaacagtttggtgcgacccctatgaattgctct 60		
548 tgcctttaggttaggaacacccctggaacagtttggtgcgacccctatgaattgctct 607		
61 acgggggacccccccctgtgtaaaatgctcttcgtacatagttgatgtaattgctctt 120		
608 acgggggacccccccctgtgtaaaatgctcttcgtacatagttgatgtaattgctctt 667		
121 cccagccttcttctcagaagcgcctcgaagtggtgtaaggaacagcgtggaagc 180		
668 cccagccttcttctcagaagcgcctcgaagtggtgtaaggaacagcgtggaagc 727		
181 agctcggagagcctactcaggaagagagactcgaagtccatctgcttccaaagt 240		
728 agctcggagagcctactcaggaagagagactcgaagtccatctgcttccaaagt 787		
241 gagattcagctatgttagagccacgcgttcaggaacactcgaagcgtggaagagac 300		
788 gagattcagctatgttagagccacgcgttcaggaacactcgaagcgtggaagagac 847		
301 ctatctcgaactttgacacacacagcgtggtgaagtcaaggaatctccacttga 360		
848 ctatctcgaactttgacacacacagcgtggtgaagtcaaggaatctccacttga 907		

QY 361 tccatgcatcccaacgtttaagctgagcgccactcccgactcggggtgagagccgaaaga 420
|||||
Db 908 tccatgcatcccaacgtttaagctgagcgccactcccgactcggggtgagagccgaaaga 967
QY 421 ctgagaatcccttaagcttcgctccatcgccctgttcccttacttacaataatcc 480
|||||
Db 968 ctgagaatcccttaagcttcgctccatcgccctgttcccttacttacaataatcc 1027
QY 481 caagcgatgataagacgtcttaagaacgtcgaaccccaatagaccttaccctacc 540
|||||
Db 1028 caagcgatgataagacgtcttaagaacgtcgaaccccaatagaccttaccctacc 1087
QY 541 tggctgattatgacccctgatacaggtgtcactgttaaatagacccgaggtgtctcc 600
|||||
Db 1088 tggctgattatgacccctgatacaggtgtcactgttaaatagacccgaggtgtctcc 1147
QY 601 agagggacccctggtggtcgaactgcaattctcctccgattgattaaacccgtgttaa 660
|||||
Db 1148 agagggacccctggtggtcgaactgcaattctcctccgattgattaaacccgtgttaa 1207
QY 661 agcacactcccaactagtcgtaagttatggtgtctatgtctgcccagcgacagagaa 720
|||||
Db 1208 agcacactcccaactagtcgtaagttatggtgtctatgtctgcccagcgacagagaa 1267
QY 721 gagaataactggtggtgtcttgaggaaactctctgtatagagatgaggtcgctacactc 780
|||||
Db 1268 gagaataactggtggtgtcttgaggaaactctctgtatagagatgaggtcgctacactc 1327
QY 781 aagcatgagactgagaaatgacgcatctctccagagacgggtaaatctccctgttc 840
|||||
Db 1328 aagcatgagactgagaaatgacgcatctctccagagacgggtaaatctccctgttc 1387
QY 841 aatccggcccgaggcaagtaaaaaatgataaactatataagaataagctcgctcccca 900
|||||
Db 1388 aatccggcccgaggcaagtaaaaaatgataaactatataagaataagctcgctcccca 1447
QY 901 tcaactatgattatctaaagataaagtttcaactgaaagaaagaaagaaatatttcaa 960
|||||
Db 1448 tcaactatgattatctaaagataaagtttcaactgaaagaaagaaagaaatatttcaa 1507
QY 961 aagtgataaattggtatgagctggtggaatagtttttaataataatggtcggtgagagag 1020
|||||
Db 1508 aagtgataaattggtatgagctggtggaatagtttttaataataatggtcggtgagagag 1567
QY 1021 tccactttaaccattcgctctatagatagagagcggtggaacccctgtggaatggga 1080
|||||
Db 1568 tccactttaaccattcgctctatagatagagagcggtggaacccctgtggaatggga 1627
QY 1081 cccgataaagtaagctggtcgaacaggggccccggcccttgtagcccaacgcaataactgccc 1140
|||||
Db 1628 cccgataaagtaagctggtcgaacaggggccccggcccttgtagcccaacgcaataactgccc 1687
QY 1141 gttccccaattaacctcgctggtggtcgtgacataacacagccgctgtgcaacagttaccac 1200
|||||
Db 1688 gttccccaattaacctcgctggtggtcgtgacataacacagccgctgtgcaacagttaccac 1747
QY 1201 ggtattgattcctacacacagccttagaaactcccaaggtgttctgtttaagaaagagag 1260
|||||
Db 1748 ggtattgattcctacacacagccttagaaactcccaaggtgttctgtttaagaaagagag 1807
QY 1261 agactcttcagttcatccagggagcttccaaagcatcaactccacgacactgagtc 1320
|||||
Db 1808 agactcttcagttcatccagggagcttccaaagcatcaactccacgacactgagtc 1867
QY 1321 actctctctgtgtgctgtgtctatcctcagggcctctcttatttgagggagatgctaa 1380
|||||
Db 1868 actctctctgtgtgctgtgtctatcctcagggcctctcttatttgagggagatgctaa 1927
QY 1381 gaaagaaatcaatgtgacaaagagatagaaatcaatgtataggggtgtccgaat 1440
|||||
Db 1928 gaaagaaatcaatgtgacaaagagatagaaatcaatgtataggggtgtccgaat 1987

QY 1441 aagcttaccctcaactgaaattcccggaaggagcatgcataggaagatcccccacc 1500
|||||
Db 1988 aagcttaccctcaactgaaattcccggaaggagcatgcataggaagatcccccacc 2047
QY 1501 caacaaacacttctgctatagtaactgtgtgttataatgagcagcctcaagaaatcaattta 1560
|||||
Db 2048 caacaaacacttctgctatagtaactgtgtgttataatgagcagcctcaagaaatcaattta 2107
QY 1561 gtaacgtgttataaagaggtgtggtgagcaatgagtttcaacccctgtgttccacc 1620
|||||
Db 2108 gtaacgtgttataaagaggtgtggtgagcaatgagtttcaacccctgtgttccacc 2167
QY 1621 tcaacttcaacaaatcccaagattgtgtgtatgctatgctccgataacccgacccaaaga 1680
|||||
Db 2168 tcaacttcaacaaatcccaagattgtgtgtatgctatgctccgataacccgaggtgtac 2227
QY 1681 taccatccgtgaggaatgttctctgtatgataatgtctctgataacccgacccaaaga 1740
|||||
Db 2228 taccatccgtgaggaatgttctctgtatgataatgtctctgataacccgacccaaaga 2287
QY 1741 gaaacccgtatcccttacccttagctgtatgctcgtatgagagcgccgtgtgctaga 1800
|||||
Db 2288 gaaacccgtatcccttacccttagctgtatgctcgtatgagagcgccgtgtgctaga 2347
QY 1801 acagggacagctgcccctgatacagagacacagcagctagagaaagagctgtgtagcta 1860
|||||
Db 2348 acagggacagctgcccctgatacagagacacagcagctagagaaagagctgtgtagcta 2407
QY 1861 catggtggtcgtgacagaaagatctccgagccttaaaaggaatctgtttagaacaactagaagag 1920
|||||
Db 2408 catggtggtcgtgacagaaagatctccgagccttaaaaggaatctgtttagaacaactagaagag 2467
QY 1921 tccctgactctctgtctgtaagtgtcttcaagaaacggagggatagatctgtctt 1980
|||||
Db 2468 tccctgactctctgtctgtaagtgtcttcaagaaacggagggatagatctgtctt 2527
QY 1981 ctaagagaaggtgtgtatgtgacagccttaaaaggaatgtgtctctatgataatcac 2040
|||||
Db 2528 ctaagagaaggtgtgtatgtgacagccttaaaaggaatgtgtctctatgataatcac 2587
QY 2041 tcaagagacatcagagacatccatgaaacaaagtctagaaaaaagtttagagagcgctogaag 2100
|||||
Db 2588 tcaagagacatcagagacatccatgaaacaaagtctagaaaaaagtttagagagcgctogaag 2647
QY 2101 gaaagagagctgacacaggggtgtgttgaagaggtgttcaacaggtctcctgtgagacc 2160
|||||
Db 2648 gaaagagagctgacacaggggtgtgttgaagaggtgttcaacaggtctcctgtgagacc 2707
QY 2161 accctgactctgtctcgaaggggccccctagtagctcctcttacttaacgttggg 2220
|||||
Db 2708 accctgactctgtctcgaaggggccccctagtagctcctcttacttaacgttggg 2767
QY 2221 cctgtcttaataatagttgtgtgtcctgtgttagagaaagatgtgtgagctgagacc 2280
|||||
Db 2768 cctgtcttaataatagttgtgtgtcctgtgttagagaaagatgtgtgagctgagacc 2827
QY 2281 atgtgacttaagcaacggttaccagagcctctgtgacaaagaaactgacacttagacc 2340
|||||
Db 2828 atgtgacttaagcaacggttaccagagcctctgtgacaaagaaactgacacttagacc 2887
QY 2341 tcccaagtctaaagttagaactttaaacaagaagaagtgtgggaatgaaagatga 2400
|||||
Db 2888 tcccaagtctaaagttagaactttaaacaagaagaagtgtgggaatgaaagatga 2947
QY 2401 tgcacacttaacccctcccaagacccaagaaagttaataaagctctaaatgtcccccgaat 2460
|||||
Db 2948 tgcacacttaacccctcccaagacccaagaaagttaataaagctctaaatgtcccccgaat 3007
QY 2461 acagacccctgtgtgtgctgacgtaaaatagatgaaggtcaacacttctatgttccagggc 2520
|||||
Db 3008 ccagacccctgtgtgtgctgacgtaaaatagatgaaggtcaacacttctatgttccagggc 3067
QY 2521 ctgtatacctgtgctaaagtaagataacaggaatgtgtgactaatgccttatctgatt 2580

```

|||||
Db 3068 ctgcatctccgccaagtaagataacagaaatagttgatactatctgattc 3127
QY 2581 ctgtaaactactgagcacaatagaatattgatacattgaaagccctagtgacta 2640
QY 3128 ctgtaaacacgactgacacataagaaatattgatacattgaaagccctagtgacta 3187
QY 2641 tctcaactgcaatctgcaactctgccaaggaagccacagagatgagcgactccgagacta 2700
Db 3188 tctcaactgcaatctgcaactctgccaaggaagccacagagatgagcgactccgagacta 3247
QY 2701 tttaaaatgattgattcacaagagcgagcggtctcgatattttaaattgattgattcaga 2819
Db 3248 tttaaaatgattgattcacaagagcgagcggtctcgatattttaaattgattgattcaga 3307
QY 2761 gagcgagcggtc-tcgatattttaaattgattgattgagcgagcggtctgattgaga 2879
Db 3308 gagcgagcggtc-tcgatattttaaattgattgattgagcgagcggtctgattgaga 3367
QY 2820 acccataaaagctgtccgattccgactcgagcgagcgagctcctacccctcgatggt 2879
Db 3368 acccataaaagctgtccgattccgactcgagcgagcgagctcctacccctcgatggt 3427
QY 2880 gtaagactgtgagcgagcgagcggtctgagaaataatcctctgtgattgagc 2934
Db 3428 gtaagactgtgagcgagcgagcggtctgagaaataatcctctgtgattgagc 3482

```

RESULT 2

US-08-766-528-2
Sequence 2, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESS: LAHAYE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MGP-038CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-5941

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 7333 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-766-528-2

Query Match

63.7%; Score 1882.8; DB 4; Length 7333;

Best Local Similarity 80.1%; Pred. No. 0; Mismatches 402; Indels 195; Gaps 9;
Matches 2409; Conservative 0;

```

QY 1 tgccttttaagtgatgaagaaacacccctgagcagttgggtctacccctatgattgctc 60
Db 4462 Tgccttttaagtgatgaagaaacacccctgagcagttgggtctacccctatgattgctc 4521
QY 61 acggggagaccccccgcttgtagaattgctctgtacatgctgattgtgctctt 120
Db 4522 ACGGGGAGACCCCCCGCTTGAGAAATGCTTGCATAGTGTCTGATGTCTGCTGCTT 4581
QY 121 cccagccttctctcctagctcagagcgctgagtgagtgagtgagtgagtgagtgag 180
Db 4582 CCCAGCCTTCTCTCCTAGCTCAGAGCGCTGAGTGAGTGAGTGAGTGAGTGAG 4641
QY 181 agctccgagagcgctcactacagagagagagagagagagagagagagagagagag 240
Db 4642 AGCTCCGAGAGCGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4698
QY 241 gagatcagctcactgtagagcgacccgctgagagagagagagagagagagagagag 300
Db 4699 GAGATCAGCTCAGTATGTTAGAGCGACCGCTGAGAGAGAGAGAGAGAGAGAGAG 4758
QY 301 ctatctcgtactcttctacacacacacacacacacacacacacacacacacacacac 360
Db 4759 CTTATCTGTACTTTTACACACACACACACACACACACACACACACACACACAC 4803
QY 361 tcatatcatccacgtaagctgagcgacacccctcagctgagtgagagcgagaaaga 420
Db 4804 ----- 4803
QY 421 ctgagatccccccttaagctcgtcctcactcagcctggtctcactcactaataatccc 480
Db 4804 -----GATCCCTTAAGCTCGCTCCATCGCGGGTCTTACTGTGCAATACTCT 4859
QY 481 caagcagatgaagaacgcctctatagacagctcgaaccccatgacctatccctacc 540
Db 4860 CAAGTTAAGTGAACGCTTGTGAGACGCCGAGACCCCATTAACCTTATCTCTCAGC 4919
QY 541 tgcctgattatgacctgatacagtgatgctgacgtgaaataagcactcagagtgctcct 600
Db 4920 TGGTTACTTACTGACTCCGCTACGATTAATTAATTAACACACACTCAAGGGAGGCTCC 4979
QY 601 aagagacacgtggtgagctgacatcttcgctcagatgattgattaccccgctgta 658
Db 4980 TTGGGACCTGTGGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5039
QY 659 -----aaagcagacccctcactagctcgtgagtgatggtctatgtgctcag 711
Db 5040 GACCGAGCCACACCCCGATGATCTCGTGTGAGGCTTACGTTTACGTTTACGTTTAC 5099
QY 712 acagagaaagaaataactgtgaggtgtcctgaggaatcctctgtgagagatgagctgc 771
Db 5100 CCAATTAATGAAGATATTGTGGAATTCCTGAGATTTCTTTTGCACAAAGAGAGCTGC 5159
QY 772 gtacactccaagatgagagactggaatgagcagctcctcctcagagacgggttaattc 831
Db 5160 ATAACTCTTAATGATGGAATTGGAATGGAATGGAATGGAATGGAATGGAATGGA 842
QY 832 tccctgtg-----ca 842
Db 5220 Tccttttgaacatctcctacagcttgaatgaattgaattgaattgaattgaattga 5279
QY 843 ttcgagccggagcagatgacaaatgataactataaagataagagctgctccactc 902
Db 5280 GATTGCGAAGCGGGTACAAAGATG-TAGCAATTAAGCAAAATTAATTAATTAATTA 9338
QY 903 agactgattatctaaagaaagttcactgaaagaaagaaagaaagaaagaaagaaagaa 962
Db 5339 AGACTGATTACTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5398
QY 963 gtgagtaaatgtagagctgaggaatgatttta-----ataatgagggggggagcag 1019

```


Db 5399 GTGGTAAATGCTATATCTTGGGAAATAGTACTATGAGGCTCTGGAGAAAGAAAG 5458
Qy 1020 gtcaacttaaccatctgccttaggtagagagcgagagacacccctgtggaatggg 1079
Db 5459 ATCTGTTCTGACTATTCGCTCAGAAATAGAAATCAGATGAGAACTCCGGTGTATAGG 5518
Qy 1080 accgataaagtaactgtgcaagagggcccgccctgagagccacgaactatgc 1139
Db 5519 ACCAAATAGGTTTGGCCGAAACAGGACCTCAATCCAAAG----- 5560
Qy 1140 gggcccaacttaaccctgcgcctgacatacaacagccgcttagcaagatcac 1199
Db 5561 -----ACAGAGGCCATCTCTTACCCCTCTGATTAATACAACTC 5602
Qy 1200 tggatgattcctaccacaacgcttagaactcccgagtgctccgttaagacagaca 1259
Db 5603 TGGATCAGTCCCACTGA-----GCCCTAACATCTATTAATAACAGGCGC 5647
Qy 1260 gggactctcaagctcagagagcttcacagccatcaactccacgacccctgagc 1319
Db 5648 GAAACTTTTATGCTCCTCAGAGAGCTTTTCAAGCTTAACTCAACGACTCCAGAGC 5707
Qy 1320 caactctctgttgctgtctatcctcagggcctccttataatgaagggatgctaa 1379
Db 5708 TACCTCTTCTGTGGCTTGGCTTACCTTCGGGCCACTTACTATAGGGAATGGCTAG 5767
Qy 1380 agaaagaaatccaatgtagcaagaagagacatagaatcaatgtagaggggctccgaa 1439
Db 5768 AGAGAGAAATTCATATGACAAAGACATAGACCAATGATACATAGGAGATGCCAA 5827
Qy 1440 taagcttaaccctcaagaggttcggaggaagggagatgtagagaaagcccccctc 1499
Db 5828 TAAGCTTACCTTACTGAGGTTTCTGAAAGAGGACCTGATGAGGATGGTTCCTCC 5887
Qy 1500 ccaacaacacctgtgctatagtagctgtgtttagagcagccctagaacacagatctt 1559
Db 5888 CCACCAACACCTTTGTAACCACTCACTGAGGCTTTATGCAACTCTGAGAGTCAATATCT 5947
Qy 1560 agtactgtgttatacaagagtggtgagcagatgaactgttaacccctgtgttcac 1619
Db 5948 GGTACTGTGTATATGACAGGAGTGGGATGATGATGATGATGATGATGATGATGATG 1679
Qy 1620 ctcaactcaacaaatccaagatgtgtgtagtgcagagtgagagagccgctggagtag 1739
Db 6008 CTGTGTTTCAACCAACTAAAGACTTGTGCTATGATGATGATGATGATGATGATGATG 1799
Qy 1680 ctcaactcctgaggaagtggtgtgtatgtagatgtagatgtagatgtagatgtagatg 1859
Db 6068 CTACTATCCGAAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 6127
Qy 1740 agaacccgataccctacccctagctgtatgtagtgcagagtgagagagccgctggagtag 1799
Db 6128 AGAGCCCATATCTCTGACACTAGCTGTATGATGATGATGATGATGATGATGATGATG 6187
Qy 1800 aacaggaagcagctgcctgtagacagagcagcagcagcagcagcagcagcagcagcag 6187
Db 6188 AACAGAGAGGCTGCTCCTTACAGAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 6247
Qy 1860 acatgagcagtagcagagagatctcgagccttaagagagtgtagcagcagcagtaga 1919
Db 6248 ACATCGAATGTTAAGGAGATCTCAAGCCCTAGAAAATCTGTAGTAACTGAGAGGA 6307
Qy 1920 gtccctcaactcttctggaagtggttcaagaacccgagggagtagatgtagtctgtt 1979
Db 6308 ATCCCTAACCCTCTTATCTGAAGTGTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 6367
Qy 1980 tctaaagagaagtggtttagcagccttaagagaagagatgtgtctatgtagatga 2039
Db 6368 TCTAAAGAGAGAGGATTATGATGATGATGATGATGATGATGATGATGATGATGATG 6427
Qy 2040 ctcaagagcagtagagagcagtagcagagcagtagaagagagagagagagagagagagag 2099
Db 6428 CTCAG 6487

Qy 2100 ggaagagagcagcagcag 2159
Db 6488 GGAAG 6547
Qy 2160 caccctgtctgtctgtgag 2219
Db 6548 CACCTGCTTGTCTGTCTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6607
Qy 2220 gctgtgttaataatag 6667
Db 6608 GCTTGTCTTAAATAGGTTGT 6727
Qy 2240 ttccagcttcaagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 2339
Db 6668 CATGTAAGTAAAG 6787
Qy 2400 atgcaacttaacccctccag 2459
Db 6788 ATGCAACTTAACCTTCCAG 6847
Qy 2460 taagaacccctgt 2519
Db 6848 TCCAG 6907
Qy 2520 cctgtctatccctgt 2579
Db 6908 CCGTGTATCTTGT 6967
Qy 2580 tctgttaaacagtagctgt 7027
Db 6968 TCTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7087
Qy 2700 attttaaatgtagtgcag 2759
Db 7088 ATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7147
Qy 2760 ggaagcag 2819
Db 7148 GAG 7207
Qy 2820 acccctaataagcgt 2879
Db 7208 ACCCATTAAG 7267
Qy 2880 gtag 2939
Db 7268 GTAG 7327
Qy 2940 aaaaaa 2945
Db 7328 AAAAAA 7333

RESULT 3
US-09-111-085-1
Sequence 1, Application US/09111085
Patent No. 6100034
GENERAL INFORMATION:
APPLICANT: Stoye, Jonathan P
TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
FILE REFERENCE: 4238/75168
CURRENT APPLICATION NUMBER: US/09/111,085

; CURRENT FILING DATE: 1998-07-07
 ; EARLIER APPLICATION NUMBER: GB 9710154.7
 ; EARLIER FILING DATE: 1997-05-16
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2462
 ; TYPE: DNA
 ; ORGANISM: Porcine retrovirus
 US-09-111-085-1

Query Match 52.9%; Score 1563.4; DB 3; Length 2462;
 Best Local Similarity 79.5%; Pred. No. 0;
 Matches 2000; Conservative 0; Mismatches 401; Indels 116; Gaps 8;

QY 151 tcgagtggtgaggaacagcgtggaagcagctccggagggccctactcagagagagag 210
 Db 1 tcgagtggtgaggaacagcgtggaagcagctccggagggccctactcagagagag 210
 QY 211 acttcaagttccacatcgctccaagttggagattcagttctatgttagagccacogt 270
 Db 58 acttcaagttccacatcgctccaagttggagattcagttctatgttagagccacogt 270
 QY 271 caggaaacctcgactcgttggagggacattatctcgtacttctgacacacacacgg 330
 Db 118 caggaaacctcgactcgttggagggacattatctcgtacttctgacacacacacgg 330
 QY 331 ctgtgaaagtcgaaggaaatctccacctggatccatccatccatccatccatccatcc 390
 Db 178 ctgtgaaagtcgaaggaaatctccacctggatccatccatccatccatccatccatcc 390
 QY 391 ctccagctcgggtgagagcgcgaagaaagactgagaatcccttaagcttccctccatc 450
 Db 238 ctccagctcgggtgagagcgcgaagaaagactgagaatcccttaagcttccctccatc 450
 QY 451 gcttggttcttactctcaataactccacagcagtagtaaaagcctttagacagc 510
 Db 298 gcttggttcttactctcaataactccacagcagtagtaaaagcctttagacagc 510
 QY 511 tcgaaccccatagacattatccctacatcgctgatttgaacctgatacgggttc 570
 Db 358 ccgaactcccatagacattatccctacatcgctgatttgaacctgatacgggttc 570
 QY 571 actgtaataagcactgaggtgtctcctagagcagcactggtgacctgaactgattc 630
 Db 418 aatattacagcactaaaggagggcctccttgggaacctggtgacctgaactgattc 630
 QY 631 tgcctccgattgattaaaccccgctgta-----aaagcacacacccccaactgacct 684
 Db 478 tgcctccgattgattaaaccccgctgta-----aaagcacacacccccaactgacct 684
 QY 685 agttatgggttctattgtgcccag--cacagagaagaagaataactgtggggttct 741
 Db 538 gcttacgggttctattgtgcccagcccccacaaataatgagaataattgtggaaatcct 597
 QY 742 ggggaatcctcttagagatggagctgcgtccactccacagatgagactgaagaatgg 801
 Db 598 caggattctctttagaagaatggagctgcataacttctaattgaattgggaatgg 657
 QY 802 ccgactctctcaggaacccggtaaaattctccttctgtaattcc----- 846
 Db 658 ccgactctctcaggaacccggtaaaattctccttctgtaattcc----- 846
 QY 847 -----ggcccgggcgaagtgaaggattggcaacaggggacaaaagatgt- 776
 Db 718 caatttaattatggccatggagatggaaggattggcaacaggggacaaaagatgt- 776
 QY 873 actatataaagataagagctgctcccatcagacttagattatataaagataagttcac 932
 Db 777 acgaataaagataaagctgtcattcgttagacacttagattacttaaaataagttcac 936
 QY 933 tgaagaaggaacacaggaataattcaaaagtgagataaattggtatgagctggggaatag 992

Db 837 tgaagaaggaacacagaataattcaaaagtgggtaaattggtatattcttgggaatagt 896
 QY 993 ttttta---ataatatggcgggagcagcaggttccactttaaccattcgccttagataga 1049
 Db 897 gtactatggaggtctctgggagaagaagagatctgttctgactattcgcctcagaataga 956
 QY 1050 gacggggacagaaacccctctggcaatgggaccccgataaagtactgctgaacagggccc 1109
 Db 957 aactcagatggaaacctccggttgcctatagagaccaaataaagggttggccgaacagacc 1016
 QY 1110 ccggccctgagccacccgcataaacttgcggtgcccccaattaaacctcgtcgggctga 1169
 Db 1017 tccaatccaagaaca-----gaggccatc 1040
 QY 1170 catacacagccgcctcagcaacagataccactggattgattcctaccacacgcctagaaa 1229
 Db 1041 tccataacctctgattacaatacaacctctggatcagtcagtcaccaactga----- 1088
 QY 1230 ctccccaggttctcgtttaagacagacagagactcttcagttcattccacagggacttt 1289
 Db 1089 ---gcctaacatcactattaaacagggcgaaaacttttagctcattccacagggacttt 1145
 QY 1290 ccaagcctcaactccacagccctgatccacttcttcttcttcttcttcttcttcttctc 1349
 Db 1146 tcaagcttcaactccacagccctcagggctcactcttcttcttcttcttcttcttctt 1205
 QY 1350 agggcctccttattatgagggatggctaaagaaaagaaatcaattgtaacaaagagca 1409
 Db 1206 gggccaccttactatgaggaatggctagagagaggaataatcaattgtaacaaagagca 1265
 QY 1410 tagaaatcaatgtacatgggggtcccgaaataagcttaacctcaactgaagtttccgggaa 1469
 Db 1266 tagagacaaatgtacatgggggtcccgaaataagcttaacctcaactgaagtttccgggaa 1325
 QY 1470 ggggacatgcataggaagagctcccccattccacacacaccccttgaacacacagcagc 1529
 Db 1326 aggaacctgcatagggatggttccccattccacacacaccccttgaacacacagcagc 1385
 QY 1530 ttatgagcagcctcagaaaaatcagttatttagtacctgttataacagagtggtggcctg 1589
 Db 1386 ctttaactgaacctgagagtcatactctgttacctggttatgacaggtggtggcctg 1445
 QY 1590 caactctgggttaacccctgtgttccacctcagtttcaacaaatccaaagatttctg 1649
 Db 1446 taactctgattaaacccctgtgttccacctggttccacctggttccacctgaagactttg 1505
 QY 1650 tgcctggttccaaatcgttcccccgagtgtaacctacctcgtgaggaagtggtccttgatga 1709
 Db 1506 cgttatggtccaaatgtcccccggtgtactactatccgaaaaagcagctcttgatga 1565
 QY 1710 atatgactatcggtataaccac 1769
 Db 1566 atatgactatgataataaccgcaaaagagagcccaatacctcgaacttagctgaat 1625
 QY 1770 gctcggattagggagcggcgttggcgttaggaacagggagcagctccctgatacagagacc 1829
 Db 1626 gctcggattggagtggtcgtcagggcgtgggaacaggaacggtcctccttaaccagagacc 1685
 QY 1830 acagcagatagagaagaagactgtgtgagctacatcgcccatcgacagaagaatctccgagc 1889
 Db 1686 gcaacagctgagagaagaagacttagtaacctacatcgaaattgtaacggaagatctcccaagc 1745
 QY 1890 ctttaagaggtctgttagcaaccttagaagagctccctgactcttcttctgctgaagtgtct 1949
 Db 1746 cctagaataatctgtcagtaacctggaggaatccctcctcctcctcctcctcctcctcct 1805
 QY 1950 acagaacccgggggattagatctgctgttcttaagaagaagggtgggttattgtgacgctt 2009
 Db 1806 acagaagaagaagggtttagatctgttatttcttaaaagaaggaggttattgttagcctt 1865
 QY 2010 aaaaagaagaatgctgcttctctatgtagatcactcaggaagccatcagagaactccatgaacaa 2069

Query Match	41.78;	Score 1231.8;	DB 4;	Length 8060;
Best Local Similarity	69.8%;	Pred. No. 0;		
Matches 1873;	Conservative	0;	Mismatches 697;	Indels 113; Gaps 11;
QY	279	ctcgaagactcgggtgaaggagaccttactctgctacttttgaccacacacacagcgctgtgaaa	338	
DB	1	CTCGAGACTCGGTGGGAAGGGCCCTTATCTGTA	CTTGTGCTACTTTTGACCAACACACGGCTGTGAAA	60
QY	339	gtcgaagaactctcacactgagatccatcgcgcatccacagctaaagctgagcgccacactcccgac	398	
DB	61	GTCGAAGGAAGTAATCCACCTGGATCCATGTCATGCCACGCTTAAGCGGGCGCACCTCCGCAT	120	
QY	399	tcggggtggaagcgcaaaaagactgagaa	tccctcttaagcttcgcgtccatcgctcggtt	458
DB	121	TGGGGGTGGAAGCGCAAAAGACTGAAATCC	CCCTTAAGCTTCGCCTCCATCGCGTGTT	180
QY	459	ccttactctacaataaactccccgggccagtagtaaaacgccttatagacagctcgaaacc	518	
DB	181	CCTTACTCTGTCAATAAACCTCTCAGACTAATG	TATGCTATGGCATAGGAGACAGCGCTGATC	240
QY	519	ccatagacatttaccctactcactgcgtgattatgacccgtgatacaggggtgtcactgtaaa	578	
DB	241	CCATAAACCTTATCTCTCACCCTGGTAA	TTACTGACTCCGGCACAGGTATTAATATCAA	300
QY	579	tagcactcagaggtgtgttccttagaggaacctggdggccctgaaactgattctgcctccg	638	
DB	301	CAACACTCAAGGGAGGCTCCTTTAGGAACCTGGTGGCC	TATGATCTATACGTTTGCCTCAG	360
QY	639	attgattaaaccccgctgttaaaagcacacactcccaactagtcgtagtctatgggttcta	698	
DB	361	ATCAGTTATTC---	CTAGTCTGACCTCACCCCCAGATATCCTCCATGCTCACGGATTTTA	417
QY	699	ttgctgcccagg---	cacagagaagaagaaatactgtggggggtttctggggaatactctctg	755

Db 1429 TCCTGTGATGTCGACCTTGAGTGGCAGAGGTGTAGGACAGGACAGCTGCCCTGCT 1488
QY 1821 cacagaccacagcagctagagaaagacttggtagctacatcgcgccatgacagaaga 1880
Db 1489 CACGGGACCACAGCAGCTAGAAACAGGACTTAGTACCTACATCGAATTTGAACAGA 1548
QY 1881 tctccgagcctaaagagctgttagcaacctagagagctccctgaacttctgtctga 1940
Db 1549 TCTCCAAGCCTAGAAAAATCTCTCAGTACCTGGAGGAATCCCTAACCTCTTATCTGA 1608
QY 1941 agtggcttcaagaacacaggggattagatctgttcttaagagaggtgggttatg 2000
Db 1609 AGTAGTCTACAGAAATAGAGAGGTTAGATTATTTCTTAAAGAGGAGGATTATG 1668
QY 2001 tgcagccttaaaagaagaatgttgcctctatagatacactcagagagccatcagagactc 2060
Db 1669 TGTAGCCTTGAAGAGGAATGCTGTTTTATGTGGATCATTCAGGGGCCATCAGAGACTC 1728
QY 2061 catgaacagcttagaaaaaagttagagagggcgctgaaggggaagagagggctgaccagg 2120
Db 1729 CATGAACAACCTTAGAGAAAGTTGGAGAAGCTCGAAGGCTCGAAGGAAAGAACTACTCAAGG 1788
QY 2121 gtgtttgaagatggttcaacagctctccttggatgacacccctgtcttctgtctgac 2180
Db 1789 GTGTTGAGGATGGTTCACAGGTCCTCTTGGTTGGCTACCCCTACTTTCTGCTTTAAC 1848
QY 2181 gggggccctagtagtctgtcctgttacttacagttggcctgtgcttaataatagtt 2240
Db 1849 AGGACCTTAATAGCTCTCTCTCTTACTCACAGTTGGGCCATGTATTATAACAAGTT 1908
QY 2241 tgttgcctttagaagaacagtagtgagtcggtccagatcatgttacttaggcaacagta 2300
Db 1909 AATTGCCCTTCATTAGAGAACCAATTAAGTGCAGTCCAGATCATGTACTTAGACAACAGTA 1968
QY 2301 ccaagccctcttgagccagagagaaactgacctctgacctccagttccagttctaaattagaa 2360
Db 1969 CCAAGCCCGCTAGC---AGGGAAGCTGGCCGCTACTCTTACAGGTTCTTAAGATTAGAA 2025
QY 2361 tatttaacaagaagaagtggaatgaagagatgaataaaggaatgaataaaggaatgaataa 2420
Db 2026 CTATTAAACAGAGAGAGTGGGAATGAAGGATGAATAAATCAACCTTAAGCTAATGAGA 2085
QY 2421 accaggaagttaataaaagctctaaatgccccgaattacagacccctgctggctgcca 2480
Db 2086 AGCTTAAATTTGTTCTGAATTCAGAGTTTGTTCCTTATAGGTAAAGATTAGGTTT 2145
QY 2481 gtaaataggtagaaggttcacacttcctattgttccagggcctgctatcctggcct-aagt 2539
Db 2146 GCTGTTTAAATATGCGCAAGTAAATAGGCCCTGAGTACATGTCTCTAGGCATGAAC 2205
QY 2540 aagataacaggaatgagttgac---taategcttatctggtattctgttaaaactgactg 2595
Db 2206 TCTTGAACACTATTGTAGATAACAAGAAAGGGAGTTCTTAAGTCTTTTAGCTTCG 2265
QY 2596 gcaccatagaagaattgattacacattgacagccctagtgacctatctcaactgcaact 2655
Db 2266 TAAACCTGGTTGCGCCATAAGATGTGAATTTGATACACATATCTTGGTGACAAAT 2325
QY 2656 gtcaactgcccagagccacacagatcgagacccctcgagctatttttaaaatgattgg 2715
Db 2326 GTCTCCCCACCCGAAACATGCGCAATGTGTAACCTCTTAAACAATTTAAATTAATGG 2385
QY 2716 tccacgagagcgggcctctcgatttttaaaatgattggttccatgagcgcggtctctg 2775
Db 2386 TCCACGAAGCGCGGCTCTCGAAGTTTAAATGTACTGGTTGT-----G 2430
QY 2776 atattttaaagtattggtttagacacagggcttggtagaaccaccataaaagctgt 2835
Db 2431 ATATTGTAATGATTGGTTGTAAAGCGCGGGCTTTGCTGTGTAACCCCATAAAGCTGT 2490
QY 2836 ccgattccgcactcgggcgccagtcctctacccctcggtggtgtagactgtggccc 2895
Db 2491 CCGGACTCCACACTCGGGGCCCGCAGTCTCTACCCCTCGGTGGTGTAGACTGTGGGCC 2550

QY 2896 cagcgcccttggaataaaaaatctcttctgtgttgcataaaa 2938
Db 2551 CAGCGCGCTTGGGAATAAAAAATCTCTTGTGTTCATCAAGA 2593

RESULT 6

US-09-376-781-25
; Sequence 25, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376.781
; EARLIER FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Contiguous
; OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

Query Match 32.9% Score 973.8; DB 4; Length 2000;
Best Local Similarity 70.2%; Pred. No. 1.2e-305;
Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;

QY 332 caccacacgctgtgaagtcgaaggaatctccacctggatccatgcacccacgttaagc 381
Db 1 caccacacgctgtgaagtcgaaggaatctccacctggatccatgcacccacgttaagc 60
QY 382 tggcgccacccctccgactcggggtgagagccgaaagagactggaatcccttaagcttc 441
Db 61 cggcgccacccctccgactcggggtgagagccgaaagagactggaatcccttaagcttc 120
QY 442 gctccatcgctgggttctcttacttacttaacataactccccagccagtagtaaacgctt 501
Db 121 gctccatcgactggttcttacttctgtaataacctctcagactaatggtatgcata 180
QY 502 atagacagctcgacccccaatagacccctttatcccttacctggtgattgacctgat 561
Db 181 ggaacagcctgacccccaatagacccctttatcccttacctggtgattgacctgat 240
QY 562 acgggtgctactgtaaatagcactcagaggtgtgtcctcctagagccactggtggcctgaa 621
Db 241 acaggtattaatatccacagcgtcgagggaggtcctcttagaaccttggtggcctgat 300
QY 622 ctgacttctgcctcgattgatttaaccccgctgtttaaagcacacccctcccaactagtc 681
Db 301 ctatagctgcctcgactcagctcctcctcctcctcctcctcctcctcctcctcctcct 357
QY 682 cgtagttatgggttcttctgctgcccagg---cacagagaaagagaaatactgtgggggt 738
Db 358 cgtgcttaccgattttatgttggccagggaccaccaataatggaaacactatggaaat 417
QY 739 tctgggaatcctctgttaggagtgagctgcctcactccacagctggagactggaaa 798
Db 418 cctagagatttcttttacaacaaatggagctgtgtaacctctaatgatggaatcgaaa 477
QY 799 tggcagatctctccagacccgggtaaaattctcttctgtaattccggtggcgccgcaag 858
Db 478 tggcacaacctctcgaggttagggtaagctttcttatgtcaaccccccaataaaccacgg 537

QY 859 tacaaaatgatgaactataaagataagagctgtctcccatcagacttagattacta 918
 Db 538 acctggaacacgtacacaggtaccattctgggtgtttccctcagacactgattatct 597
 QY 919 aagataagttccactgaaagaaagaaagaaatattcaaaagtggaataatggtatg 978
 Db 598 aaaaataagttccacccgaaagaaagaaagaaatattcaaaagtggaataatggtatg 657
 QY 979 agctgggaatagtttttta---ataatattggcggggagcaggtccactttaaccatt 1035
 Db 658 tccctggggaataatattataacaggttggaacagacacacaggtccactttaaccatt 717
 QY 1036 cgccttagatagacgaggggacagaaacccctgtggcaatggggagcccgataaagtactg 1095
 Db 718 cgaactaaata---agcagctagagctcccaatggctatadagacccaatcaggtctta 774
 QY 1096 gctgaacagggggcccccgcctgagccacccgacataacttgcggtgcccataaacc 1155
 Db 775 acgggtcaaaagaaaccccccaccccgagccacccctct-----810
 QY 1156 tcgctgcggcctgacataacacagcgcctagcaacagtagtaacacactgaggtatttactacc 1215
 Db 811 -----gatataacttctaattagaccccaact 837
 QY 1216 aaacgcctagaaactcccgaggtgttctgttaagacagacagagactcttcagtcctc 1275
 Db 838 gagtctaacagcagcagc-----taaacggggacaaaacttttttagctc 882
 QY 1276 atccaggagcttccaaagccatcaactccacccgacactgtatgcactcttctgttgg 1335
 Db 883 atccaggagcttccaaagccatcaactccacccgacactgtatgcactcttctgttgg 942
 QY 1336 ctttctctatcctcagggcctcctctattataggggagtgctaaagaaagaaattcaat 1395
 Db 943 ctttctctatcctcagggcctcctctattataggggagtgctaaagaaagaaattcaat 1002
 QY 1396 gtacacaaagacagataaataatgacatgaggggtcccgaaataaagcttaccctact 1455
 Db 1003 gtacacaaagacagataaataatgacatgaggggtcccgaaataaagcttaccctact 1062
 QY 1456 gaagtttccgggaggggacatgcataagaaagctcccccacccacacacacactttgc 1515
 Db 1063 gagggttcttgaaagacactctgataaaaggttccccccacccacacacacactttac 1122
 QY 1516 tatagtactgttttatagcagcgcctcagaaaaacagttatttagtactgggtataac 1575
 Db 1123 aaccacactgaagcctttaaatacaaacctctgagagcaatactgtgtacctgttatgac 1182
 QY 1576 aggtgtgggacatgcaatactgggttaacccctgtgtttccacactcagtttcaaccaa 1635
 Db 1183 aggtgtgggacatgcaatactgggttaacccctgtgtttccacactcagtttcaaccaa 1242
 QY 1636 tccaaagattgtgtcatgttccaaatcgtcccccagtgtaactaccactcctgaggaa 1695
 Db 1243 actaaagactttacattatggtcccaattgttccccagtgataactatcccaagaaa 1302
 QY 1696 gtggtccttgatgaatgatctatcggtataacccgacccaaagagacacccgtatccct 1755
 Db 1303 acaattctcgatgaatgattacaggaacacatcgacaaagaaacccatccctcctg 1362
 QY 1756 accctagctgtaattgctcgtattagggagcggcgtgtggcgtaggaacagggagcagctgc 1815
 Db 1363 acactcgcagtaattgctcgtattagggagcggcgtgtggcgtaggaacagggagcagct 1422
 QY 1816 ctgatacagggacacacagcagctagagaaagggactgtgtgagctacatcgggccatgaca 1875
 Db 1423 ttagttaacagacccacagcagctagaaacagggacttagtaacctacacataaattgtaaca 1482
 QY 1876 gaagatctccagccttaaaagagctctgttagcaacctagaagagtcctcctgactctttg 1935
 Db 1483 ggaaatctccagccttaaaagagctctgttagcaacctagaagagtcctcctgactctttg 1542

QY 1936 tctgaagtggttctacagaaccggagggattagatctgttcttaagagaaggtggg 1995
 Db 1543 tctgaagtggttctacagaataaaaggggttagatttatttcttaaaaaaagaaga 1602
 QY 1996 ttatgtcagccttaaaagaaagatgttcttctatgtatcactcagagccatcaga 2055
 Db 1603 ttatgttagccttaaaagaaagatgttcttctatgtatcactcagagccatcaga 1662
 QY 2056 gactcccatgaacagcttagaaaaagtttagagagcgtcgaaagggaaagagaggtgac 2115
 Db 1663 gactcccatgaacagcttaaaaaagtttagagagcgtcgaaagggaaagagaggtgac 1722
 QY 2116 -caggggtgttgaagagatggttcaacaggtctcttctgtatgacacacgtgttcttcgc 2174
 Db 1723 tcaagatggtttaaagaggttcaacaggtctcttctgtatgacacacgtgttcttcgc 1782
 QY 2175 tctgaaggggccccttagtctctctcttcttacttactacagttgggoccttcttaataa 2234
 Db 1783 ttaacagggaccccttaagtagtccctcccttactcacagttgggoccttcttaataa 1842
 QY 2235 taggtgttgccttctttagaagaacaggtgagtgagtgagtgacagtgacttaggca 2294
 Db 1843 caagttaattgcttattagaaacacgaataagtcagtcagtcagtcagtcagtcagtc 1902
 QY 2295 acagtaacaaagccttctgacccaggaagaaactgactcttagccttccagttcttaaga 2354
 Db 1903 acagtaacaaagccttctgacccaggaagaaactgactcttagccttccagttcttaaga 1959
 QY 2355 ttagaactattaaacaaagaaaggtggggaatgaaggat 2395
 Db 1960 ttaaaactattaaacaaagaaaggtggggaatgaaggat 2000

RESULT 7
 US-09-376-781-30
 ; Sequence 30, Application US/09376781
 ; Patent No. 6261806
 ; GENERAL INFORMATION:
 ; APPLICANT: Banerjee, Papi T.
 ; APPLICANT: Patience, Clive
 ; APPLICANT: Andersson, Goran K.
 ; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
 ; TITLE OF INVENTION: Use
 ; FILE REFERENCE: 61750-267
 ; CURRENT APPLICATION NUMBER: US/09376781
 ; CURRENT FILING DATE: 1999-08-18
 ; EARLIER APPLICATION NUMBER: 60/097,015
 ; EARLIER FILING DATE: 1998-08-18
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 30
 ; LENGTH: 2000
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
 ; OTHER INFORMATION: consensus sequence of PERV-D.
 US-09-376-781-30

Query Match 32.9%; Score 973.8; DB 4; Length 2000;
 Best Local Similarity 70.2%; Pred. No. 1.2e-305;
 Matches 1461; Conservative 0; Mismatches 532; Indels 88; Gaps 8;

QY 322 caccacagcgtgtgaaagtcgaaggaatctccacccgtgacccatgacccacgttaagc 381
 Db 1 caccacagcgtgtgaaagtcgaaggaatctccacccgtgacccatgacccacgttaagc 60
 QY 382 tggcgccacccctccagcactcgggtggagccgaaagactgagaatcccttaagcttc 441
 Db 61 cggegccacccctccagcactcgggtggagccgaaagactgagaatcccttaagcttc 120

QY	442	gctccatcgctgggttctctacttaacataactcccagcgcagtagtaaagcctt	500
Db	121		
QY	502	atagacagctgaaccoccatagacatttatccoctacotcggctgatttgacctgat	180
Db	181		
QY	562	acgggtgtcaacttaaaTAGCACTCGAGGTGTTGTCCTCCTAGACGCACCTGGTGGCTGA	621
Db	241	acaggtaataataccacagcgtcagggagagcttccttlagaacaactcggtygctgat	300
QY	622	ctgcaattctcctccogattgattaacccgcgtttaaagcacacactccccaacctagt	681
Db	301	ctaatagtctcctccagtcagtcattc---ctagtctgaacctcaaccacgatctctc	357
QY	682	cgtagttatgggttctattgtctccacgg---cacagagaagagaaaaactctgtgggggt	738
Db	358	cgtcttacggattttatgtttgccagagaccaccaataatggaaaacactatggaat	411
QY	739	tctgggaatcctctctagagatggagctgcgtcacctccaacgatggagactggaaa	798
Db	418	octagatatcttttacaaaacatggagctgtgaacctclaatgataggaaatcoga	477
QY	799	tggccgctctctccaggaccgggtaaattctcttctgaattccggccggccggaag	858
Db	478	tggccacactctctgcgataaggttaagctttcttatgttaaccccaataaacaccgg	537
QY	859	tacaaaatgatgaactataaagataagagctgtctcccactcaacttagattatata	918
Db	538	acctggaaaacgtcalacaggttaccattctgggtgttttccctcagactagattatct	597
QY	919	aagataagtttcactgaaaaagaaacagggaaaaatatccaaggtagataatgggatg	978
Db	598	aaaataagtttcaccgaaaaaaaacaagaataatcctaaatggataaatlygtatg	657
QY	979	agctggggaatagtttttta---ataatatggcggggagcaggttccactttaaccatt	1035
Db	658	tcttggggaataatacatatacaggttgggacagacacccagctccattctaaaccatc	717
QY	1036	cgcttagatagagagcgggacagaaacccctctggcaatgggacccgataaagtactg	1095
Db	718	cgacttaaaata---agccagctagagctccaatggctataggagcaataacggttcta	774
QY	1096	gctgaacaggggcccctggcctggcgcacgcgataacttgcgggtgcgcccaataacc	1155
Db	775	acgggttcaaagaagcccacccacccgaccatctct-----	810
QY	1156	tgcgtcgccctgacataacacagcgcgctagcaacgtagtaaccagtgattgattctacc	1215
Db	811	-----gataacttctaaatttagaccacct	837
QY	1216	aacagccttagaaactccccagggtgtctctgtaagacagacagagactcttcagctctc	1275
Db	838	gagtctaacagcacac-----taaaacggggacaaaacttttagctctc	882
QY	1276	atccaggagctttccaaagccatcaactccaacgacctgatgccattctcttgttgg	1335
Db	883	atccaggagcttttcaagctccLaactccaacgactccagagcctctcttgttgg	942
QY	1336	ctttgtctatcctcaggctctcttatatgaggggatggctaaagaagaataatcaat	1395
Db	943	ctttgttaactcgggccaccttactaLaagaaatggctaaagagagaaaaattcaat	1002
QY	1396	gtgaccaaagagcatagaaaLcaatgtfacatgggggtcccgaaataagcttacctcact	1455
Db	1003	gtgacaaaaaacatagagaccaatgtacatggggaLccccaaataagcttaccttact	1062
QY	1456	gaagtttccgggaaggagacatgcatgagaaaagctcccccaLccccacacacotttgc	1515
Db	1063	gaggtttctggaaggacactcgataaaaaaaggttcccccatccccacacacottttac	1122
QY	1516	tatagctactgtggtttatgagcaggcctcagaaaaatcagtatatttagtacctgggtataac	1575

RESULTS

RESOLUTION 8
US-09-376-781-24

03 03-370-781-24
: Sequence 24. Application IIS/09376781

; Patent No. 6261806

; GENERAL INFORMATION:

APPLICANT: Banerjee, Papiya T.

APPLICANT: Patient, Clive

; APPLICANT: Andersson, Goran K.

;
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

; Patent No. 6261806

; TITLE OF INVENTION: Use

; FILE REFERENCE: 61750-267

; CURRENT APPLICATION NUMBER: US/09/376,781

Query Match	24.08;	Score 710.4;	DB 4;	Length 1493;	
Best Local Similarity	70.79;	Pred. No. 3.4e-220;			
Matches 1069;	Conservative	0;	Mismatches 361;	Indels 82;	Gaps 6;
QY	890	gctgctcccatccagacttagattatctaaagataagttctcactgaaaaggaacacagg	949		
Db	50	ggtgtttccctcagacactagattatcttaaaataagttccacgaaaaaaaacaag	109		
QY	950	aaaatactcaaaagtggaataaaggatgaactgggaatagttttta---ataatag	1006		
Db	110	aaaatactctaaaatggataatggta tgcctggggaataataattatacaggttggg	169		
QY	1007	gcggggagcaggggtccactttaaccattcgcttagtagagagcgggacagaacccc	1066		
Db	170	acagacaaccaggtccattctaaacctcagacttaaaaa---agccagctagagcctc	226		
QY	1067	ctgtggcaatgggacccgataaaagtactggctgaacaggggcccgcgcctggagccac	1126		
Db	227	caatggctatagacggaatcaggtcttaacgggtcaagaaccccaccccaggacct	286		
QY	1127	cgataaacttgcgggtgcccgaattaaactcgctgcggctgacataaacacagccccta	1186		
Db	287	ccctgatataac-----ttctaaattagaccacacagctgagt	322		
QY	1187	gcaacagtagaccactgattgattctctacacacagcctagaaaactcccagggtgtccgt	1246		
Db	323	ctaacagcagac-----	335		
QY	1247	ttaagacaggaacagagactctcagttctcatccagggagctttccagagcatcaactcca	1306		
Db	336	-taaaacgaggacaaaacttttagtctcatccagggagctttccagctcttaactcca	394		
QY	1307	ccgacctgtagtgcacttcttctgtgtgtgtgtctctatcctcagggccctctattatg	1366		
Db	395	cgaactcagaggtactcctctctgtgtgtgtgtgttaactctggccacactactata	454		
QY	1367	aggggatggctaaagaagaaaattcaattgacacaaagagcagacagaaatcaatgtacat	1426		
Db	455	agaaaa tggctaaagaagaaaaattcoattgtacaaaaaaacatagagaccaatgtacat	514		
QY	1427	gggggtcccgaaataagctttacccctcaactgaagttcttcgggaaggggacatgcataagaa	1486		
Db	515	ggggatcccaaaa taagctttacccctactgaggtgtcttggaaagagacacctgcataaaa	574		
QY	1487	aagctcccccattcccacacaccttggtagtagtactgtgtgttatgagcagggcctcag	1546		
Db	575	aggttccccattcccccaacacctttacacacacacgaagcctttaatcaaacctctg	634		
QY	1547	aaaatcagatttagtacctgggtataacacagtggtgggcatgcataactctgggttaaccc	1606		
Db	635	agagtcaatatctgttacctgggttagacaggtgggtggcatgtaatactgatttaaccc	694		
QY	1607	cctgtgtttccactcagctctcaacacatccaaagatttgggtcattgggtccaaatcg	1666		
Db	695	cttgtgtttccactctgggttttcaacacacacaaagactttacattatgggtccaaattg	754		
QY	1667	tcccccgagtgataccactcctgaggaagtggtccttgatgaatatgactactcgttata	1726		

```

1  RESULT          9
2  US-09-011-745-7
3  ; Sequence 7, Application US/09011745
4  ; Patent No. 6165715
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Collins, Mary KL
7  ; APPLICANT: Weiss, Robin A
8  ; APPLICANT: Takeuchi, Yasuhiro
9  ; APPLICANT: Cosset, Francois-Loic
10 ; TITLE OF INVENTION: Expression systems
11 ; FILE REFERENCE: 09/011,745
12 ; CURRENT APPLICATION NUMBER: US/09/011,745
13 ; CURRENT FILING DATE: 1998-06-22
14 ; EARLIER APPLICATION NUMBER: PCI/GB96/02061
15 ; EARLIER FILING DATE: 1996-08-23
16 ; EARLIER APPLICATION NUMBER: GB95117263.1
17 ; EARLIER FILING DATE: 1995-08-23
18 ; NUMBER OF SEQ ID NOS: 29
19 ; SOFTWARE: Patentin Ver. 2.0
20 ; SEQ ID NO 7
21 ; LENGTH: 6312
22 ; TYPE: DNA

```

	ORGANISM:	Artificial Sequence	
:	FEATURE:		
:	OTHER INFORMATION:	Description of Artificial Sequence: Portion of	
:	OTHER INFORMATION:	construct	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4058)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4059)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4061)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4246)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4247)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4248)	
:	OTHER INFORMATION:	n is any nucleotide	
:	FEATURE:		
:	NAME/KEY:	misc_feature	
:	LOCATION:	(4249)	
:	OTHER INFORMATION:	n is any nucleotide	
:	US-09-011-745-7		
	Query Match	15.4%; Score 455; DB 4; Length 6312;	
	Best Local Similarity	53.0%; Pred. No. 8.5e-137;	
	Matches 1261; Conservative	0; Mismatches 1030; Indels	88; Gaps
Qy	2	gccttttaggggttaggaacaccctcgagcagtgttggtggtgacctatgaattgctcta	61
Db	1208	gcTgtagggccagaatacccccTggcgggttTggtttaactccttatgaattctcta	1267
Qy	62	cgggggaccccccccgcttggtgaaattgctctgatcatgctgctgcttc	121
Db	1268	TgagdgaccaccccccaactTggtCtGgagaaactTgggtcccgatagattct	1327
Qy	122	ccagcctttgtctcTaggtcaaggcgcgtcgagtggctgagccaagcgtggaacca	181
Db	1328	ccctgcttatttaactcaacttaaaggcttagaaaattgaaggaccacaatctgggacca	1387
Qy	182	gctccggggagcgcctactcaggagagagacttgcaagttccaatcgcttccaagtTg	241
Db	1388	gatcaaagaggTgataaagccTg---gtacggtaaacaatccctcacccgttccagTcgg	1444
Qy	242	agattcagctctatgttagacgcacccgtTgcaggaaacctcgagactcggTggaaggacc	301
Db	1445	ggatcaagTgctTgtcagacgcctatcgaccagcagcctTgagctcggtTgaaaagccc	1504
Qy	302	ttaetctgTactTTtgaccacaccacgcctTgaagtctgaagctgaagaatatccacctgat	361
Db	1505	atacctTgTgtTgctgactTaccacccgacgcgTaaaaagctgattgTctgctcggtT	1564
Qy	362	ccatgatccccagcttaagctggcgccaacct-----ccgactcggggTggagacgga	415
Db	1565	ccatgcttctcaactcaaacctTgacccacctTcgccaccagatgagTcctTggagctTgga	1624
Qy	416	aagactgagaatcccccttaagcttgcctccatcgctggtgTtcttactct-aacaaTa	474

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: POOF, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-075-272-1

Query Match 12.6%; Score 372.2; DB 3; Length 8655;
Best Local Similarity 62.1%; Pred. No. 8.3e-110;
Matches 663; Conservative 0; Mismatches 383; Indels 21; Gaps 4;

QY 1250 agacaggacagagactcttcagtcacatccagggagagcttccagccatcaaatccaccg 1309
DB 6706 ATACGAAACCGCTGTGCTAGCTAGTTCAGGGAGCCCTTTTATGTTTAAATAGGACTA 6765
QY 1310 accctgagccactctcttctgtgtgttctatccctcagggcctctcttatgatgg 1369
DB 6766 ATCCTAATGACTCAATCATGTTGGTTATGCTATGCTCTTAACCCCTTATATGAAG 6825
QY 1370 ggaatggaagaagaataatcaatgtgacaaagagcatagaaataatgatcatggg 1429
DB 6826 GAATTGCTCAGACTAGAACTTACATAATATTACTTACAGATCAT--TCTCAATGCTTTGGG 6882
QY 1430 ggtccgaaataagcttaccctcactgaagttccggaggaagggacatgcataagaaag 1489
DB 6883 GAGAGAACAGGAAGTTGACTCTGACGAGGTTTCAGGAATGGGCTTTGTTAGGTGAGG 6942
QY 1490 ctcccccattccacacacacctttgctatagtagtctgtgtttatgagcagcctcagaaa 1549
DB 6943 TGCCCCAGGATAAATGGCACCTCTGTACACGACTCAAAATATCCGACCTAACAAAGGTG 7002
QY 1550 atcagtagtagtacctggttataacaggtgtgtgggcatgaactactgggttaacccccc 1609
DB 7003 GTCAATATCTAGTGTCTCCCATAGACAGATATGGCTTGCATACAGAGTCTCACTCCTT 7062
QY 1610 gtgttccacatcagctcttcacaaatccaagatttgtgtcatgtgtccaaatctcc 1669
DB 7063 GTATATCTATGTCTGTTTCAATAGCTCCAAAGATTCTGTATTTTATGCTCAGCTTATTC 7122
QY 1670 cccgagtgtagtaccatcctgaggaagtggtccttgatgaatatgactatcggtataacc 1729
DB 7123 CTAGACTCTGTATCATGATGATAGTATCTATTTCTAGACAAATTTCAACATCGG--GTCC 7179
QY 1730 gacaaaagaagaaacccgtatccctaccctagctgtgaatctcggatt---agggcgg 1786
DB 7180 GCTGAAAGAGAAACCAATATCTTAACTGTAGCAGTCTTTTAGGATTTGGAGTGGGAGTGCCG 7239
QY 1787 ccgttggcgtggaacagggacagctgcctgatacagagggaccacagcagctagagaaaag 1846
DB 7240 CAGCTGGGGTAGGTACTGAACCGCTGCTTAATCCACACCCCGATCTTTGAG----7295
QY 1847 gacttggtagcatcgtcgccatgacagagaagatctccgagccttaagaggtctgtta 1906
DB 7296 -----GAATTAGGTACAGCTATGGTACTGTACTCTCAGAGCTATAGAACACTCTATAA 7347

US-08-716-351A-5
Sequence 5, Application US/08716351A
Patent No. 6033905
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
TITLE OF INVENTION: Retroviral Vectors
NUMBER OF SEQUENCES: 5
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,351A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: PCT/US94/03784
APPLICATION NUMBER: PCT/US94/03784
FILING DATE: 06-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 15280-128-1PC
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..10970
OTHER INFORMATION: /standard_name= "p537 retroviral"
OTHER INFORMATION: vector"
US-08-716-351A-5

Query Match 12.1%; Score 357.8; DB 3; Length 10970;

us-09-171-553b-9.feb22std.rni

Mon Feb 25 07:44:33 2002


```

1311 GGAGGCGCTCACCCTCAACGATGAGACTGGAATGCGCCATCTCTCTC 1360
151  GlnAspArgValLysPheSerPheValAsnSerGlyProGlyLysTyrLys 167
1361 CAGGACCGGGTAAATCTCTCTTGTCAATCCGGCCCGGCGCACTAATA 1410
167  smethelysLeuTyrLysAspLysSerCysSerProSerAspLeuAspT 184
1411 ACTGATGAACCTATATATAAGATAGAGCTGCTCCCATCTAGACTTAT 1460
184  yTLeuLysLLeSerPheThrGlyLysGlyLysGlnGluAsnLLeGlnLys 200
1461 ATCTAAGATAGATTCTCACTGAAAAAGAAAAACAGAAAAATATTCAAA 1510
201  TrpIleAsnGlyMetSerTyrGlyLysValPheTyrLysTyrGlyGly 217
1511 TGGATTAATGGTATGAGCTGGGGAATGTTTTTAAATATGGGGGGG 1560
217  ValAlaGlySerThrLeuThrIleArgLeuArgLLeGluThrGlyLys 234
1561 AGCAGAGGCTCCTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1610
234  roProValAlaMetGlyProAspLysValLLeuAlaGlnGlnGlyPro 250
1611 CCCCTGTGGCAGTGGGACCCGATTAAGTACTGGCTGAACAGGCGCC 1660
251  AlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLeu 267
1661 GCCCTGGAGCCACCCATTAATCTGGCGGCTGCCCAATTAACCTG 1710
267  gProAspIleThrGlnProProSerAsnSerThrThrGlyLeuLeuPro 284
1711 GCTTACATTAACACAGCGCTTACGACAGGTAACCTGATGATGATCT 1760
284  hrAsnThrProArgAsnSerProGlyValProValLysThrGlyLys 300
1761 CCAACAGCCCTAGAAACTCCCGAGGTCTCTGTTAAGACAGGACAG 1810
301  LeuPheSerLeuLLeGlnGlyAlaPheGlnAlaIleAsnSerThrAsp 317
1811 CTCTTCACTTCATCCAGGAGCTTCCAGCCATCACTCCACCGACCC 1860
317  AspAlaThrSerSerCysTrpLeuGlyLeuSerGlyProProTyr 334
1861 TGATGCCACTTCTCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCT 1910
334  yTGLuGlyMetAlaLysGluArgLysPheAsnValThrLysGlnHis 350
1911 ATGAGGGATGGCTAAAGAAAGAAATTCATGTGACCAAGACATAGA 1960
351  AsnGlnCysThrTyrGlySerArgAsnLysLeuThrLeuThrGluVal 367
1961 AATCAATATGATGAGGGGTCCGAAATAGCTTACCTCACTGAAGTT 2010
367  rGlyLysGlyThrCysLLeGlyLysAlaProProSerHisLHisLLeu 384
2011 CGGGAAGGAGCATGATGAGAAAGCTCCCATCCACCAACACCTTT 2060
384  yTyrSerThrValValTyrGlnGlnAlaSerGluAsnGlnTyrLeuVal 400
2061 GCTATAGACTGCTGTTATATGACAGGCTCAGAAATCAGATTTTAT 2110
401  ProGlyTyrAsnArgTyrTrpAlaCysAsnThrGlyLeuThrProCys 417
2111 CCTGTTATTAACAGTGGTGGCATGCAATCTGGGTTAACCCCTGTGT 2160
417  LSerThrSerValPheAsnGlnSerLysAspLeuCysValMetValGln 434
2161 TTCACCTCAGCTTCAACCAATCAAGATTTCTGTCTCATGTGTCCAA 2210
434  LeValProArgValTyrTyrHisProGlnGluValLLeuAspGluTyr 450
|||||
2211 TCGTCCCCGAGTGTACTACATCTGAGGAAGTGTCTTGATGATAT 2260
451  AspTyrArgTyrAsnArgProLysArgGluProValSerLeuThrLeu 467
2261 GACTATCGGTATTAACCCAGAAAAAGAAACCCGTATCCCTTACCT 2310
467  aValMetLeuGlyLeuGlyThrAlaValGlyValGlyTyrGlyThrAla 484
2311 TGTATCTCTGGATTAGGAGCGCGCTGGCTGATGAAACAGGAGACCT 2360
484  ILeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGlnLysLLeu 500
2361 CCTGATACAGAGACACAGCAGCTAGAGAAAGACTGGTGTGATACAT 2410
501  AlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAsn 517
2411 GCGGCATGACAGAAAGATCTCCAGGCTTAGAGAGAGTCTGTAGCA 2460
517  uGluGluSerLeuThrSerLeuSerGluValLLeuGlnAsnArgArg 534
2461 AGAAGAGTCCCTGACTTCTTGTGTAAGTGTCTTACACAAACGAG 2510
534  LLeuAspLeuLeuPheLeuArgGlyGlyLysCysAlaAlaLeuLys 550
2511 GATTAGATCTGCTCTTCTTAAGAAAGTGGTATGTGACGCTTAAAA 2560
551  GluGluCysCysPheTyrValAspHisSerGlyAlaIleArgAspSer 567
2561 GAAAGAAATGTTGCTTATGATGATCACTCAGAGCATCAGACATCCAT 2610
567  LAsnLysLeuArgLysLysLeuGluArgArgArgGluArgGluAla 584
2611 GACCAAGCTTAAAGAAAGTGTAGAGAGGCTCGAAAGGAAAGAGG 2660
584  spGlnGlyTyrPheGlnGlyTyrPheAsnArgSerProTrpMetThr 600
2661 ACCAGGAGTGTGTTGAAGATGTTTCAACAGGCTCTCTTGATGAC 2710
601  LeuLeuSerAlaLeuThrGlyProLeuValLLeuLeuLeuLeuLeu 617
2711 CTGCTTCTCTCTGACGGGACCCCTAGTATGCTCTCTCTGTTACT 2760
617  rValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArg 634
2761 AGTTGGGCTTGTCTTAATTAAGTTGTGTGCTTGTGTAGAGAACG 2810
634  aLserAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 650
2811 TGATGTCAGTCCGATCATGTTACTTACGCAACAGTACCAAGCCT 2860
651  SerGlnGlyLysThrAspLeu 657
2861 AGCCAAGAGAAACTGACCTC 2881

seq_name: /cgn2_6/plodata/2/ina/6A_COMB.seq:US-09-111-085-1
seq_documentation_block:
; Sequence 1, Application US/09111085
; Patent No. 610034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2462

```

```

; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-1

```

```

alignment_scores:
    Quality: 2631.50      Length: 679
    Ratio: 4.468          Gaps: 9
    Percent Similarity: 86.745    Percent Identity: 73.049

```

```

alignment_block:

```

```

US-09-171-553B-10 x US-09-111-085-1 ..

```

```

Align seg 1/1 to: US-09-111-085-1 from: 1 to: 2462

```

```

1 MethSProthrLeuSerTrpArgHisLeuProthrarGlyGlyLeu 17
  |||||||..... |||||||..... |||||||.....
211 ATGCATCCAGCTTAAGCCGCCACCTCCGATTCGGGGTGAAGGCC 260
  |||||||..... |||||||..... |||||||.....
17 oLysArgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeu 34
  |||||||..... |||||||..... |||||||.....
261 GAAAGAGCTGAAGATCCCTTAACCTTCGCTCCATCGCGTTCCTTA 310
  |||||||..... |||||||..... |||||||.....
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSer 50
  |||||||..... |||||||..... |||||||.....
311 CTGTGTCATTAAGTCTCAAGTAAAGTAAAGCCCTTGAGACAGCCG 360
  |||||||..... |||||||..... |||||||.....
51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAsp 67
  |||||||..... |||||||..... |||||||.....
361 AACTCCCTTAAGCCCTTAATCTCTACCCGCTTACTACTCCGCTAC 410
  |||||||..... |||||||..... |||||||.....
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
  |||||||..... |||||||..... |||||||.....
411 AGGATTTAATTTTAAACAGACTCAAGGAGGAGCTCCCTTGAGGAGCT 460
  |||||||..... |||||||..... |||||||.....
84 rProGlyLeuHisPheCysLeuArgLeuIleAsnProAlaVal..... 98
  |||||||..... |||||||..... |||||||.....
461 GGCCTGAATTAATATGCTGCTCGATCAGATGATCCCGTCAATGAC 510
  |||||||..... |||||||..... |||||||.....
99 LysSerThrProProAsnLeuValArgSerTrpGlyPheTrpCysCys 115
  |||||||..... |||||||..... |||||||.....
511 CAGGCGACACCCCGGATGATCTCGGTACCGGTTTACCGTTTCC 560
  |||||||..... |||||||..... |||||||.....
115 oGly...ThrGlyLysGlyLysTrpCysGlySerGlyLysSerPheC 131
  |||||||..... |||||||..... |||||||.....
561 AGGACCCCAATTAATGAGAAATATTGTGAAATCTCAGCATTTCTTT 610
  |||||||..... |||||||..... |||||||.....
131 ysaArgArgTrpSerCysValThrSerAsnAspGlyAspTrpLysTrp 147
  |||||||..... |||||||..... |||||||.....
611 GCAAGCANTGAGCTGCATACCTCTAATGATGGGAATTGGAAATGGCCA 660
  |||||||..... |||||||..... |||||||.....
148 IleSerLeuGlnAspArgValLysPheSerPheValAsnSer..... 161
  |||||||..... |||||||..... |||||||.....
661 GTCTCTCAGACAGACAGAGTAAGTACTCTTTGTTAAACAATCCTACC 710
  |||||||..... |||||||..... |||||||.....
162 .....GlyProGlyLysTrpLys.....MetMetL 170
  |||||||..... |||||||..... |||||||.....
711 TTATATCATTAATTATTATGAGCATGGAGATGGAAGAGATTGGCAAC 760
  |||||||..... |||||||..... |||||||.....
170 ysaLeuTrpLysAspLys.....SerCysSerProSerAsp 181
  |||||||..... |||||||..... |||||||.....
761 GGGTACAAAGATGTACGAATTAAGCAATTAACCTGTCATTCGTTAGAC 810
  |||||||..... |||||||..... |||||||.....
182 LeuAspTrpLeuLysIleSerPheThrGlyLysGlyLysGlnGluAsn 198
  |||||||..... |||||||..... |||||||.....
811 CTAGATTCTTAATAAATAGTTTCACTGAAAAAGCAAAACAAAGAAATAT 860
  |||||||..... |||||||..... |||||||.....
198 eGlnTrpTrpIleAsnGlyMetSerTrpGlyLleValPheTrpLysTrp 215
  |||||||..... |||||||..... |||||||.....
861 TCAGAAAGTGGTAAATGATATCTTGAGGAATAGTGAC.....TATG 904
  |||||||..... |||||||..... |||||||.....
215 IyGlyGlyAla.....GlySerThrLeuThrIleArgLeuArgIle 228
  |||||||..... |||||||..... |||||||.....

```

```

905 GAGGCTGTGGAGAAAGAAAGATCTTCTGACATAATCCCTCAGATA 954
  |||||||..... |||||||..... |||||||.....
229 GluThrGlyThrGlyProProValAlaMetGlyProAspLysValLeu 245
  |||||||..... |||||||..... |||||||.....
955 GAACCTCAGATGGACCTCGGTGCTATAGACCAATTAAGGTTTGCC 1004
  |||||||..... |||||||..... |||||||.....
245 agLingLysProProAlaLeuGluProProHisAsnLeuProValPro 262
  |||||||..... |||||||..... |||||||.....
1005 CGAACAAGAGACCTCCATCCAGAACAGACG..... 1032
  |||||||..... |||||||..... |||||||.....
262 InLeuThrSerLeuArgProAspIleThrGlnProProSerAsnSer 278
  |||||||..... |||||||..... |||||||.....
1033 .....AGGCCATCTCTTAACCCCTCTGATTTACATACATCAACC 1068
  |||||||..... |||||||..... |||||||.....
279 ThrGlyLeuIleProThrAsnThrProArgAsnSerProGlyValPro 295
  |||||||..... |||||||..... |||||||.....
1069 TCTGGATCAAGTCCCACT.....GAGCCTTAACATCTCACTAT 1103
  |||||||..... |||||||..... |||||||.....
295 IlyThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAla 312
  |||||||..... |||||||..... |||||||.....
1104 TAAACAGAGGGCGGAACCTTTTATAGCTCATPCACAGGAGCTTTCAAGCTC 1153
  |||||||..... |||||||..... |||||||.....
312 LeuAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeu 328
  |||||||..... |||||||..... |||||||.....
1154 TTAATCCACAGACTCCAGAGGCTACCTCTTGTGCTTGCTTGCCTT 1203
  |||||||..... |||||||..... |||||||.....
329 SerGlyProProTrpTrpGlyGlyMetAlaLysGlyLysArgLysPhe 345
  |||||||..... |||||||..... |||||||.....
1204 TCGGGCCACCTTACTATAGGGAATGGCTAGAGGAGGGAATTCATAT 1253
  |||||||..... |||||||..... |||||||.....
345 IThrLysGlyHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeu 362
  |||||||..... |||||||..... |||||||.....
1254 GAAAGAGAACATAGAGACCAATGATGATGAGGATGCCAAATTAACCTTA 1303
  |||||||..... |||||||..... |||||||.....
362 hrLeuThrGlyValSerGlyLysGlyThrCysIleGlyLysAlaPro 378
  |||||||..... |||||||..... |||||||.....
1304 CCTTACTAGAGTCTCTGAAAGGCCACCTGCTAGAGATGAGTCCCCCA 1353
  |||||||..... |||||||..... |||||||.....
379 SerHisGlnHisLeuCysTrpSerThrValIleValTrpGlnGlnAla 395
  |||||||..... |||||||..... |||||||.....
1354 TCCACCAACACCTTTGTAACACACACTGAAAGCTTAAATGAAACCTCTGA 1403
  |||||||..... |||||||..... |||||||.....
395 uAsnGlnTrpLeuValProGlyLysAsnArgTrpTrpAlaCysAsnThr 412
  |||||||..... |||||||..... |||||||.....
1404 GATTCATATATCTGTACTGCTGATATGACAGGTGCTGGCATGTAATAC 1453
  |||||||..... |||||||..... |||||||.....
412 IyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAsp 428
  |||||||..... |||||||..... |||||||.....
1454 GATTAACCCCTTGTGTTCCACCTTGCTTTCACCAACCAATTAAGACT 1503
  |||||||..... |||||||..... |||||||.....
429 CysValMetValGlnIleValProArgValIleTrpHisProGlnGlu 445
  |||||||..... |||||||..... |||||||.....
1504 TGGGTTATGGTCCAAATGTGCCCCGGGTGATCTATATCCGAAAAAGC 1553
  |||||||..... |||||||..... |||||||.....
445 IValIleAspGlyTrpAspTrpArgTrpAsnArgProLysArgGluPro 462
  |||||||..... |||||||..... |||||||.....
1554 AGTCCTTGATGATATGACTATGATATATGCGCCAAAGAGACGCCA 1603
  |||||||..... |||||||..... |||||||.....
462 aIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaVal 478
  |||||||..... |||||||..... |||||||.....
1604 TATCCCTGACACATAGCTGATATCTCGATTTGGAGATGCTGAGGCGTG 1653
  |||||||..... |||||||..... |||||||.....
479 GlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGlySer 495
  |||||||..... |||||||..... |||||||.....
1654 GGAACAGGAAGGCTGCTCCCTTAATCAGAGACCCCAACACTGGAGAAAG 1703
  |||||||..... |||||||..... |||||||.....
495 yLeuGlyGlyLeuHisAlaAlaMetThrGlyAspLeuArgAlaLeuLys 512
  |||||||..... |||||||..... |||||||.....
1704 ACTTAGAACCTATCATCGAATTTGTAACGGAAGATCTCCAAAGCCCTG 1753
  |||||||..... |||||||..... |||||||.....
512 InSerValSerAsnLeuGlnGluSerLeuThrSerLeuSerGlyVal 528
  |||||||..... |||||||..... |||||||.....
1754 AATCTGTCACTAGCTGAGAGATCCCTTAACCTCTTATCTGAAAGGTGT 1803
  |||||||..... |||||||..... |||||||.....

```

```

529 LeuGlnAsnArgArgGlyLeuAspLeuPheLeuArgLysGlyLeu 545
1804 CTACAGAACAGAGGGGTTAGATCTGTATTCTTAAAGAGAGGGTT 1853
545 ucysAlaAlaLeuLysGlyLysCysCysPheTyrValAspHisSerGlyA 562
1854 ATGTGTAGCCTTAAAGAGAGATGCTGCTCTATAGTACCTCAGAG 1903
562 LaIeaArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArg 578
1904 CCATCAGAGACTCCATGACAGAGCTTAGAGAAAGGTAGAGAGCGCCGA 1953
579 ArgLysArgGlyLysAlaAspGlyTyrPheGlyLysGlyTyrPheAsnArg 595
1954 AGGGAAGAGAGAGGCTGACAGGGGTGCTTGAAGGATGCTTCAACAGGTC 2003
595 rProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValVal 612
2004 TCCTTGATGACACCCCTGCTTCTGCTGACGGGCCCCCTAGTGTCC 2053
612 euleuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAla 628
2054 TGCTCCTGTTACTTACAGTGGGCTTGTATATATAGTGTGTGCC 2103
629 PheValArgLysArgValSerAlaValGlnIleMetValLeuArgGln 645
2104 TTGTGTAGAGACGAGTGTGAGTGCAGTCCAGATCATGTTAGGCACAGA 2153
645 nTyrGlnGlyLeuLeuSerGlnGlyLysIleuThrAspLeu 657
2154 GTACCAAGGCTTCTGAGCCAGCAAGCAACTGACCTC 2190

seq_name: /cgn2_6/ptodata/2/ina/6B-COMB.seq:US-08-766-528-2

seq_documentation_block:
; Sequence 2, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-2

alignment_scores:
    Quality: 2537.50      Length: 658
    Ratio: 4.452          Gaps: 9
    Percent Similarity: 86.626      Percent Identity: 72.796

alignment_block:
US-09-171-553b-10 x US-08-766-528-2 ..

Align seg 1/1 to: US-08-766-528-2 from: 1 to: 7333

22 IleProLeuSerPheAlaSerIleAlaThrPheLeuThrLeuIleTh 38
4806 ATCCCTTAAAGCTTCCCTCCATCCGCTGTTCTTACTCTGTCAATAC 4855
38 rProGlnAlaSerSerLysArgLeuIleAspSerSerAsnProHisArg 55
4856 TCCTCAGATTAATGTAACCGCTTGTGACAGCCCGCAATCCCATTAAC 4905
55 rLeuSerProThrThrPheLeuIleIleAspProAspThrGlyValThrVal 71
4906 CCTTATCTCTCAGCTGTTACTTACTGACTCCGTTACAGGTAAATTA 4955
72 AsnSerThrArgGlyValAlaProArgGlyThrTrpTrpProGlyLeu 88
4956 AACACACTCAAGGGAGGAGCTCCCTGGGAGACTGTGCTGATTA 5005
88 sPheCysLeuArgLeuIleAsnProAlaVal.....LysSerThrPro 103
5006 TGCTCTGCTGATCATCAATCCCTGCTCAATGACAGCCACACCCC 5055
103 rOAsnLeuValArgSerTyrGlyPheTyrCysCysProGly...ThGlu 118
5056 CCGATGTACTCCGTCCTTACGGGTTTACGTTTCCCGCAGACCCCAAT 5105
119 LysGlnLysTyrCysGlyLysSerGlyLysSerPheCysArgArgTrp 135
5106 AATGAAGAATTTGGGAATTCACAGATTTCTTTTCAAGCAATGAG 5155
135 rCysValThrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGln 152
5156 CTGCATTAATCTTAATGATGGAATTTGGAATGCGCAGTCTCAGCAAG 5205
152 sPArgValLysPheSerPheValAsnSer..... 161
5206 ACAGATAGTTACTCTTTGTAAACAATCTTACCAGTTATATCAATTT 5255
162 .....GlyProGlyLysTyrLys.....MetLysLeuTyrLysAs 174
5256 AATTATGCGCANTGGAGATGGAAGATTTGCGAACAGCGGATCAAAAAGA 5305
174 pLys.....SerCysSerProSerAspLeuAspTyrLeu 186
5306 TGTACGAATAAGCAAAATATACCTTGTTCATCTTATGATTAATCTTAA 5355
186 ysiIleSerPheThrGlnLysGlyLysGlnLysAsnIleGlnLysTrpIle 202
5356 AATTAAGTTTCACTGAAAAAGGAAACAGAAAAATTTCAAAAGTGGGTA 5405
203 AsnGlyMetSerTrpGlyIleValPheTyrLysTyrGlyGlyAla... 218
5406 AATGATATCTTGGGGAATAGTGTAC.....TATGAGGCTCTGGGAG 5449
219 .....GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 233
5450 AAAGAAAGATCTGTTCTGACTATTCCTCAGAAATGAGAAATCAGATGG 5499
233 LuProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyPro 249

```

```

5500 AACCTCGGTGCTATAGAGCAAAATAGGTTGGCCGACAGACAGCT 5549
250 ProAlaLeuGluProProHisAsnLeuProValProGlnLeuThrSerLe 266
5550 CCATCCAGAACAGACG..... 5564
266 uArgProAspIleThrGlnProProSerAsnSerThrThrGlyLeuIleP 283
5565 .AGGCCATCTCCAAACCCCTGTGATTACATACAAACCTCTGATCAGTCC 5613
283 roThrAsnThrProArgAsnSerProGlyValProValLysThrGlyGln 299
5614 CCACT.....GAGCCTAACATCATTTAAACAGGCGG 5648
300 ArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSerThrAs 316
5649 AAACCTTTAGCCTCATCCAGGAGCTTTCAAGCTTTAACTCCACAGAC 5698
316 pProAspAlaThrSerSerCysTrpLeuGlyLeuSerGlyProProT 333
5699 TCCAGAGGCTACCTCTTCTTGGCTTGTAGCTTCGGGCCACCT 5748
333 yTrpGlnGlyMetAlaLysGluLysPheAsnValThrLysGluHis 349
5749 ACTATGAGGGAATGGCTAGAGAGGGAATTCATATGTGACAAAGAACAT 5798
350 ArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluVa 366
5799 AGAGACCATATACATGGGATCCCAAAATAGCTTACCTTACTGAGGT 5848
366 lSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisGlnHisL 383
5849 TTCTGGAAGGACACCTGCATAGGATGTTCCCATCCACCAACACC 5898
383 euCysTySerThrValAlaTyrgLuglnAlaSerGluAsnGlnTyrLeu 399
5899 TTTTGTAACCACTTAATTCACAACTCTGAAAGTCAATATCTG 5948
400 ValProGlyTyArgAsnArgTrpTrpAlaCysAsnThrGlyLeuThrProC 416
5949 GTACCTGGTTATGACAGGTGGTGGCATGTATATCTGATTAACCCCTTG 5998
416 sValSerThrSerValPheAsnGlnSerLysAspLeuGlyValMetValG 433
5999 TGTTCACACTTGGTTTCAACAACTAAAGACTTTGGCTTATGTGTC 6048
433 lInIleValProArgValTyrTrpHisProGlnGluValValLeuAspGlu 449
6049 AAATGTCCCGGGGTGACTACTATCCGAAAGAGCAGTCTGTGATGAA 6098
450 TyrAspTyArgTyArgAsnArgProLysArgGluProValSerLeuThrLe 466
6099 TATGACTATAGATATAATCGGCCAAAAGAGACCCATATCCCTGACACT 6148
466 uAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyThrGlyThra 483
6149 AGCTGTATGCTCGGATGGGATGGCTGACGCGTGGAGACAGAACAGG 6198
483 lAlaLeuIleThrGlyProGlnGlnLeuGlyLysGlyLeuGlyGluLeu 499
6199 CTGCCCTTATACAGACCGGACAGCGTGGAGAAAGAGCACTTATGAACTTA 6248
500 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSerAs 516
6249 CATCAATTTTACGAGAGATCTCCAGCCCTAAGAAAATGTGTCAGTAA 6298
516 nLeuGlnGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 533
6299 CTTGAGAGATCCCTAACCTCTTATCTGAGGTGTTCTACAGAACAGAA 6348
533 rGlyLysAspLeuPheLeuArgGlnGlyLysGlyValLeuGlyAlaIleu 549
6349 GGGGTTTATGATCTGTATTCTTAAAGAGAGGAGGTTATGTGAGCTTA 6398

```

```

550 LysGlnGluCysCysPheTyrValAspHisSerGlyAlaIleArgAsp 566
6399 AAAGAGATGCTGCTTCTATGATGATCTACAGAGCCATCAGACACTC 6448
566 lMetAsnLysLeuArgLysLysLeuGluArgArgArgGluArgGluA 583
6449 CATGAGCAAGCTTAGAGAAAGGTTAGAGAGGCGTCAAGAGGAAAGACAG 6498
583 lAspGlnGlyTrpPheGlnGlyTrpPheAsnArgSerProTyrMetThr 599
6499 CTGACAGGGGTGGTTGAAAGATGTTCAACAGAGTCTCTTGGATGAC 6548
600 ThrLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 616
6549 ACCCTGCTTGTGCTGACGGGCCCTAGTACTGCTGCTGCTTACT 6598
616 uThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluA 633
6599 TACAGTTGGGCTTGCTTATATAGTTGTTGCTTGTGTTAGACAC 6648
633 rValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
6649 GAGTGAGTGCAGTCCAGATCATGTTAGCGCAACAGTACCAAGGCTT 6698
650 LeuSerGlnGlyGluThrAspLeu 657
6699 CTGAGCCAGAGAGAACTGACCTC 6722

```

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:US-08-766-528-1

seq_documentation block:

Sequence 1, Application US/08766528

Patent No. 6190861

GENERAL INFORMATION:

APPLICANT: Jay A. Fishman

TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 60 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentL Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/766,528

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/572,645

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers

REGISTRATION NUMBER: 35,965

REFERENCE/DOCKET NUMBER: MGP-038CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8060 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-766-528-1

alignment_scores:

Quality: 2402.00 Length: 655
 Ratio: 4.312 Gaps: 9
 Percent Similarity: 85.038 Percent Identity: 68.702

alignment_block:

US-09-171-553b-10 x US-08-766-528-1 ..

Align seg 1/1 to: US-08-766-528-1 from: 1 to: 8060

```

1 MethisProthrLeuSerThrParGlyLeuProThrArgGlyGlyLys 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 ATGATCCACAGCTTAAGCGCGGCGACCTCCGATCGGGGTGGAAAGCC 135
17 OlysrArgLeuArgLeuProLeuSerPheAlaSerIleAlaThrPheLeu 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 GAAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCGGTGCTCTTA 185
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSerSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 CTCGTCTCATATACCTCTCAGACTATAGTATGCGATGAGACAGCCTG 235
51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAspThr 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 AACTCCCATTAACCTTATCTCTCAGCTGTTAAATTACTGACTCGGCGAC 285
67 rGlyAlaThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 AGTTATTTATATCAACAACACTCAAGGAGGAGCTCTTGAAGAACCTGGT 335
84 rProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 GGCCTGATCTATACGTTGGCTCAGATCATGTTATCTGATGCTGACCTCA 385
101 ThrProProAsnLeuValArgSerTrpGlyPheTrpCysCysProGlyTh 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 ...CCCCAGATATCTCCATGCTCAGCGAGTTTATGTTGCCACGAGACC 432
117 rGlyLysGlu...LysTrpCysGlyGlySerGlyLysSerPheCysArg 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 ACCAATATATGAAACACATGCGGAATCCAGACATTTCTTTGCTAAC 482
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
483 AATGGAACCTGTGTACTCTATATGATGATATGGAATGGCCAAACCTCT 532
150 LeuGlnAspArgValLysPheSerPheValAsn.....SerGlyProG 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
533 CAGCAGGATAGGGTAAGTTTCTTATGTCACACCTTATACCAAGCTCTGG 582
164 LysTrpLysMetMetLysLeuTrpLysAspLysSer.....CysSerP 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 ACAATTTATATTCCTGACCTGATTAAGAACTGAAAGCCCAAGGCTCTTC 632
179 rOserAspLeuAspTrpLeuLysIleSerPheThrGlyLysGlyLysGln 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 CTTTCAGACCTGATTAACCTTAAATTAAGTTTCTTACAGAAAGGAAACAA 682
196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTh 212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 GAAAATATCTTAATAATGGTAAATGTTATGCTTGGGAAATGGTATAT.. 730
212 rLysTrpGlyGlyGlyAla.....GlySerThrLeuThrIleArgL 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
731 ...TATGAGAGCTCGGGTAAACACACAGGCTCCATCTTAACCTATCTGCC 776
226 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 TCAAAATTAACACAG...CTGAGAGCTCCCAATGCTATAGAGCAAAATACG 823
243 ValIleuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPr 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

824 GTCTGACGGGTCAAAGACCCCAACCCCAAGACCA..... 859
259 ovalProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 .....GACCAATCTCTTA 872
276 snSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
873 ACATTAACCTCTGATCAGACCCCACTGATGCTATGACAGCAGACT..... 916
293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 .....AAATGGGGCAAACTTTTATGCTTATCAGGAGACTTT 957
309 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuc 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
958 TCAAGCTCTTAACCTCAGACCTCAGAGGCTTCTTCTGTGGCTAT 1007
326 LysLeuSerSerGlyProProTrpTrpGlyGlyMetAlaLysGluArgLys 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1008 GCTTAGCTTGGGCGCACCTTACTATGAAGGATGCTGAGAGAGGAA 1057
343 PheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySerArgAs 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1058 TTCAATGTGACAAAGAACATAGAGACCAATGCAATGGGATCCCAAAA 1107
359 nLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysA 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1108 TAAGCTTACCTTACGAGGTTCTTGGAAGAGCAGCTCATGAGAAAG 1157
376 lAbProSerHisGlnHisLeuCysTrpSerThrValValTrpGluGln 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1158 TTCCCCATCCCAACACACCTTGTACACACTCAACCTTAATCA 1207
393 AlaSerGlnAsnGlnTrpLeuValProGlyTrpAsnArgTrpTrpAlaCy 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 ACCCTGAAAGTCAATATCTGTAACCTGTTATGACAGGTGGGCGATG 1257
409 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerL 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1258 TAATACTGATTAACCCCTTGTGTTCCACCTTGTTTAAACCAAACTA 1307
426 LysAspLeuCysValMetValGlnIleValProArgValTrpTrpHisPro 442
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1308 AAGATTTTTCATTAATGCTCAAAATGTTCCCGAGATGATTAATCAATCC 1357
443 GluGluValAlaLeuAspGluTrpAspTrpArgTrpAsnArgProLysAr 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1358 GAAAACCAATCTCTGATGATGATGACTACAGAAATCATGACAAAGAG 1407
459 gGluProValSerLeuThrLeuAlaValMetLeuGlyLysGlyThAlaVal 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1408 AGAACCATATCTCTGACACTTCTGATGATGCTGAGACTGGAGTGGAGT 1457
476 alGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1458 CAGGTGAGAGACAGAACAGCTGCCCTGCTCAGGAGACACAGCGAGCTA 1507
493 GluLysGlyLeuGlyGlyLeuHisAlaAlaMetThrGluAspLeuArgL 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1508 GAAACAGGACTTGTATACCTACATCCGATTTGTAACAGAGATGCCAAGC 1557
509 alLeuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSerG 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1558 CTTAGAAATATCTGTGCTAGTAACCTGGAGAAATCCCTTAACCTCTATCTG 1607
526 LuValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgLys 542
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 AAGTAGCTTCAAGAAATGAAGAGGTTAGATTATATTTCTTAAAGAA 1657
543 GluGlyLeuCysAlaIleLeuLysGluLysCysCysPheTrpValAspH 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1658 GAGAGATTATGTGTAGCTTGAAGAGAGAAATGCTGTTTATATGTGATCA 1707

```

559 sSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLeuGluA 576
 1708 TTCAAGGGGCATCAGAGCTCCATGACAACTAGAGAAAGTTGGAGA 1757
 576 rgaTgArGArGluArGluAlaAspGlnGlyTrpPheGlnGlyTrpPhe 592
 1758 AGCGTCAGAGGAAAGAACTACTCAAGGTGTGGATGAGGATGCTTC 1807
 593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProle 609
 1808 AACAGGCTCCTTGTTGGCTACCTGCTTCTGCTTTAAACAGACCCCTT 1857
 609 uValValLeuLeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArgP 626
 1858 AATAGTCCTCCTCCTTACTCAAGTGGCCATGATATTATTAAACAAT 1907
 626 heValAlaPheValArgGluArGluValSerAlaValAlaIleMetValLeu 642
 1908 TAAATGCTTCATTAGAGAAAGAAATAGTCACTCCAGATCATGTACTT 1957
 643 ArgGlnGlnTrpGln 647
 1958 AGAACACAGTACCA 1972
 seq_name: /cgn2_6/ptdata/2/ina/6B_COMB.seq:US-08-766-528-3

seq_documentation_block:
 ; Sequence 3, Application US/08766528
 ; Patent No. 6190861
 ; GENERAL INFORMATION:
 ; APPLICANT: Jay A. Fishman
 ; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
 ; NUMBER OF SEQUENCES: 74
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/766,528
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/572,645
 ; FILING DATE: 14-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Louis Myers
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: MGP-038CP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8132 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-766-528-3

alignment_scores:
 Quality: 2395.00 Length: 655
 Ratio: 4.292 Gaps: 10

Percent Similarity: 85.191 Percent Identity: 68.855
 alignment_block:
 US-09-171-553b-10 x US-08-766-528-3
 Align seq 1/1 to: US-08-766-528-3 from: 1 to: 8132

1 MethIstProthLeuSerTrpArgHisLeuProThArgGlyGlyLeuPr 17
 5620 ATGCATCCACAGTTAAACCGCGCCACCTCCGATTCGGGGTGAAGAAC 5669
 17 OlyArgLeuArgGlyLeuSerPheAlaSerIleAlaTrpPheLeuT 34
 5670 GAAAGAGCTAAATCCCTTAAGCTTGCGCTCCATCGGTGGTTCCTTA 5719
 34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuLeuAspSer 50
 5720 CTCTGTCAATTAACCTCTCAGACTAATGTATGGCATAGACAGACCTG 5769
 51 AsnProHisArgProLeuSerProThTrpLeuIleLeuAspProAspTh 67
 5770 AACTCCCATAAACCTTATCTCTCACCTGTGTAATTACTGACTCCGCG 5819
 67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
 5820 AGGTATTATATATACACACACTCAAGGAGAGCTCTTTAGAGACCTGT 5869
 84 rPProGluLeuHisPheCysLeuArgLeuLeuAsnProAlaValLysSer 100
 5870 GGCCTGATCATACGTTTGCCCTCAGATCAAGTATTCATGCTGACCTCA 5919
 101 ThrProAsnLeuValArgSerTrpGlyPheTrpCysProGlyTh 117
 5920 ...CCCCAGATATCCCTCATCTCAGGATTTTATGTTGGCCAGGACC 5966
 117 rGlyLysGlu...LysTrpCysGlyGlySerGlyLeuSerPheCysArg 133
 5967 ACCAAATATATGGAAGAAACATTGGCGAAATCCAGACATTTCTTTGTAAAC 6016
 133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProLeuSer 149
 6017 AATGGAAGCTGTAAACCTCTTAATGATGATGATGAAATGGCCAACTCT 6066
 150 LeuGlnAspArgValLysPheSerPheValAsn...SerGlyProGly 164
 6067 CAGCAGAGTAGGGTAAGTTTCTTATGTCAACACACTATACACCTCTGG 6116
 164 LysTrpLysMetLeuLysLeuTrpLysAspLysSer...CysSerP 179
 6117 ACAATTATATACCTGACCTGATTAAGAACTGGAAGCCCAAGTGTCTC 6166
 179 rSerAspLeuAspTrpLeuLysLysSerPheThrGlyLysGlyLysGln 195
 6167 CTTGAGACCTAGTACTTAATAAATTAAGTTTCACTGAGAAAGAAACAA 6216
 196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyLysLeuPhe 212
 6217 GAAATATATCTTAATAATGGTAAATGATGCTTGGGAATGATATAT.. 6264
 212 rLysTrpGlyGlyGlyAla...GlySerThrLeuThrIleArgL 226
 6265 ...TATGAGAGCTCGGTAAACAACAGGCTCATCTTAAGTATTCGCC 6310
 226 euArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLys 242
 6311 TCATAATTAACCAAG...CTGAGCCTCAATGGCTTATAGACCAATAATACG 6357
 243 ValLeuAlaGlnGlnGlyProProAlaLeuGlnProProHisAsnLeuPr 259
 6358 GTCTTGACGGGTCAAGACCCCAACCAAGACCA..... 6393
 259 oValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSerA 276


```

576  rgatrgatrgatrgluarvgluahlaaspglnglytrp phegluslytrp phe 592
7292  AGCCTCCAAAGGAAAGGAAGAACTACTCAAGAGGTGTTTACGAGATGCTTC 7344
593  AsnAtgSerProtrpmetThrThrIleuLeuSerAlaIeuThrGlyProLe 609
7342  AACAGAGCTCTTGTTGGCTACCCACTTCTTGCTTTACAGGACCCCTT 7392
609  uValValIleuIleuIleuIeuThrValGlyProCysIleuIleAsnAcP 626
7392  AATAGTCCTCTCTCTGTTACTCTCAAGTTGGCCATGATATTATTACAAGT 7442
626  heValAlaIheValAlaGluIuArgValSerAlaValGlnIleuValIleu 642
7442  TAAATGGCTTCATTAGAGACGAATAGTCAAGTCCAGATCATGTGTTACTT 7492
643  ArgIingIinTryGln 647
7492  AGACACACAGTACCAA 7506
seq_name: /cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-376-781-25

```

```
seq.documentation block:
Sequence 25. Application US/09/36781
Patent No. 6261806
GENERAL INFORMATION:
APPLICANT: Banerjee, Papia T.
APPLICANT: Patience, Clive
APPLICANT: Andersson, Goran K.
TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Me
Patent No. 6261806
TITLE OF INVENTION: Use
FILE REFERENCE: 61750-267
CURRENT APPLICATION NUMBER: US/09/376,781
CURRENT FILING DATE: 1999-08-18
EARLIER APPLICATION NUMBER: 60/097,015
EARLIER FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 2000
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Continuous
OTHER INFORMATION: compilation of SEQ ID NOS: 18, 22, and 24.
US-09-376-781-25

alignment_scores:
Quality: 2169.00 Length: 652
Ratio: 4.032 Gaps: 11
Percent Similarity: 82.515 Percent Identity: 64.571

alignment_block:
US-09-171-553B-10 x US-09-376-781-25 ..

Align seg 1/1 to: US-09-376-781-25 from: 1 to: 2000

1 MethisPProThrLeuSerTPArgHisLeuProThrArgGlyGLuPr 17
:::|||||::::||| ||||| | ::||| :||| :|||
43 ATACATCCAGCGTTAAGCGGGGCCCACTTCGTATTACAGGGTAAGAACC 92
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
93 AAAAAGACTAAATAATCCTTAAAGCTTGCGCTCCATCACATGTTCTTA 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 olysAtgLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuT 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
93 AAAAAGACTAAATAATCCTTAAAGCTTGCGCTCCATCACATGTTCTTA 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 hrLeuThrIleThrProGlnAlaSerSerLyAsnArgLeuIleAsnSerSer 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 CTCTGTCAATAACCTCTGCAGACTAATGAATGACATGACATGAGACAACCTG 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 AsnProHisArgProLeuSerProThrTrpPheIleIleAsnProAsph 67
```


seq_name: /cgn2_6/plodata/2/ina/6B_COMB.seq.us-09-376-781-30

seq_documentation_block:

Sequence 30, Application US/09376781

Patent No. 6261806

GENERAL INFORMATION:

APPLICANT: Banerjee, Papia T.

APPLICANT: Patience, Clive

APPLICANT: Andersson, Goran K.

TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

Patent No. 6261806

TITLE OF INVENTION: Use

FILE REFERENCE: 61750-267

CURRENT APPLICATION NUMBER: US/09/376,781

CURRENT FILING DATE: 1999-08-18

EARLIER APPLICATION NUMBER: 60/097,015

EARLIER FILING DATE: 1998-08-18

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 30

LENGTH: 2000

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Nucleotide

US-09-376-781-30

alignment_scores:

Quality: 2169.00

Ratio: 4.032

Percent Similarity: 82.515

Percent Identity: 64.571

alignment_block:

US-09-171-553b-10 x US-09-376-781-30

Align seg 1/1 to: US-09-376-781-30 from: 1 to: 2000

```

1 MethisProThleuSerTrpArgHisLeuProThrArgGlyGluIrr 17
  ::::::::::::::::::::
43 ATACATCCACGTTAGCGGCGCCACCTCGATTGAGGTAAGGAAAGCC 92
17 olysArgLeuArgGlyProLeuSerPheAlaSerIleAlaTrpPheLeu 34
  ::::::::::::::::::::
93 AAAAAGACTAAATCCCTTAAGCTTCGCTCCATCAGATGTCCTTA 142
34 hrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAspSer 50
  ::::::::::::::::::::
143 CTCTGTCATTAACCTCTCAGACTAATGTATGCACATAGACAGACCTG 192
51 AsnProHisArgProLeuSerProThrTrpLeuIleIleAspProAsp 67
  ::::::::::::::::::::
193 AACCTCCATAAACCTTATCTCGACCTGTTAATTAAGTCTGACAC 242
67 rGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThrTrp 84
  ::::::::::::::::::::
243 AGGTATTAATATCCACAGCGCTCGAGGAGGAGCTCCTTAGAACTG 292
84 rPProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
  ::::::::::::::::::::
293 GGCTGATCTATATGTCTGCTCAGATCAGTCACTTCTAGTCTGACCTCA 342
101 ThrProAsnLeuValArgSerTrpGlyPheTrpCysCysProGlyThr 117
  ::::::::::::::::::::
343 ACC...CCAGATATCTCTCGCTTACGATTTTATGTTGCCAGGACC 389
117 rGlyLysGlu...LysTrpCysGlyLysArgGlyLysArgSerPheCysArg 133
  ::::::::::::::::::::
390 ACCAATTAATGAAACACTATGGAATCTAGAGATTTCTTTTACAAAC 439
133 rGTrpSerCysValThrSerAsnAspGlyAspTrpLysTrpProIleSer 149

```

```

440 AATGAGCTGTGTAACTTAATGATGAAATGAAATGCAACCTCT 489
  ::::::::::::::::::::
150 LeuGlnAspArgValLysPheSerPheValAsnSerLysProGlyLys 166
  ::::::::::::::::::::
490 CTGCAAGATAGGTAAGCTTTCTTATGTCAAC...CCCAATACCA 533
166 rLysMetLeuLysLeu...TyrLysAspLysSer...CysSerProSerA 181
  ::::::::::::::::::::
534 CCGGACCTGGAAACGTCATACAGAGTACCAATTCGGGTGTTTCCCTCAG 583
181 sPLeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGluAsn 197
  ::::::::::::::::::::
584 ACTAGATTTATCTTAAATAGTTTACCGCAAAAAAACAAGAAAT 633
198 rLeuGlnSerTrpIleAsnGlyMetSerTrpGlyLysValPheTrpLys 213
  ::::::::::::::::::::
634 ATCTTAAATGATTAATGATGTCTCTGGGGAATTAATATTAATACAGG 683
214 rTyrGlyGlyValAlaGlySerThrLeuThrIleArgLeuArgIleGlu 230
  ::::::::::::::::::::
684 TTGGGACAGACACACGCTCCATTTACACATCCGACTTAATAATA...A 730
230 hrGlyThrGluProProValAlaMetGlyProAspLysValLeuAlaGlu 246
  ::::::::::::::::::::
731 GCCAGTAGAGCTCCCAATGAGCTATAGACCGAATAGCGCTTAACGGGT 780
247 GlnGlyProProAlaLeuGluProProHisAsnLeuProValProGln 263
  ::::::::::::::::::::
781 CAAGAACCCCAACCCAGACGACCTCTGATATA... 816
263 uThrSerLeuArgProAspIleThrGlnProProSerAsnSerThrArg 280
  ::::::::::::::::::::
817 ...ACTTCTAATTAAGCCCACTAG...TTTACAGACAGCACT. 855
280 rLysLeuIleProIleAsnThrProArgAsnSerProGlyValProValLys 296
  ::::::::::::::::::::
856 .....AAA 858
297 ThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeu 313
  ::::::::::::::::::::
859 ACGGGACAAACACTTTTATGCTCATCCAGAGACTTTTACAGCTCTTA 908
313 nSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSer 330
  ::::::::::::::::::::
909 CTCACAGACTCCAGAGGCTACCTCTCTGTTGCTTCTTACTTACCTCG 958
330 rProProTrpTrpGlyGlyMetAlaLysGluArgLysPheAsnValThr 346
  ::::::::::::::::::::
959 GCCACCTTACTATAGAAATGGCTAAAGAGAAATTCAAATGATGACA 1008
347 rLysGlnHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThr 363
  ::::::::::::::::::::
1009 AAAAAGACTTAAGACCAATGATGAGGATCCCAAAATTAAGCTTACCT 1058
363 uThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerH 380
  ::::::::::::::::::::
1059 TACGAGAGTTCTGGAAGAGACACCTGCATTAATAAGTTCCCATCC 1108
380 rGlnHisLeuCysTrpSerThrValValTrpGluGlnAlaSerGluAsn 396
  ::::::::::::::::::::
1109 ACCAACACCTTTACACACACCTGACCTTAATTCAAACCTCGAAGT 1158
397 GlnTrpLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnProGlyLe 413
  ::::::::::::::::::::
1159 CAATATCTGTAACCTGTTATGACAGAGTGAGGCAATGATATGATGATT 1208
413 uThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
  ::::::::::::::::::::
1209 AACCCCTTGTTCCACCTGTTTCAACCAAACTAAAGACTTTTACA 1258
430 alMetValGlnIleValProArgValTrpTrpHisProGluGluValVal 446
  ::::::::::::::::::::

```

```

1259 TTAATGTCCTCAAAATGTTCCCGAGTATATTACTATCCCAAGAAACAAATTT 1308
447 LeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGluProValSe 463
|||||
1309 CTGATGATGATATATATTACAGGAACCATCGCAAAAGAAAGAAACCCCATATC 1358
463 rLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
|||||
1359 CCGACACTCGCAGTAATGCTCGACTCGAGATACAGATACAGGTGTAGAA 1408
480 hrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGlyValGlyLeu 496
|||||
1409 CAGGAACCTGCAGCTTGTAGTACAGGACCCAGCAGCTAGAACAGACATT 1458
497 GlyGluLeuHisAlaIleMetThrGluAspLeuArgAlaLeuGlyLeu 513
|||||
1459 AGTAACCTACATCAATTTGTACAGAAATCTCCAGACCTTAAAAAAATC 1508
513 rValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuG 530
|||||
1509 TGTCACTAACCTCGAAGAAATCCCTACCTCTATCTGAAGTATGTTCTAC 1558
530 LysAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuGlyLeu 546
|||||
1559 AGATATAAAAGGCTTAGATTATTTCTAAAAAAGAAAGATTTATGT 1608
547 AlaAlaLeuLysGluGluLysCysPheTyrValAspHisSerGlyAlaIle 563
|||||
1609 GTAGCCTTAAAGAGAAATGCTGTTTATGTATGATCATTCAGGGGCCAT 1658
563 eArgAspSerMetAsnLysLeuArgLysLysLeuGluLysArgArgArgG 580
|||||
1659 CAGAGACTCCAGTAAACACACTTAAAAAAGTTGGAGAAACGTGCAAGGG 1708
580 LysArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgSerP 596
|||||
1709 AAAAGGAACTTACTCTAAAGATGGTTTAAAGATGGTTTAAACAGGTCTCC 1758
596 oTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 613
|||||
1759 TTGGTTGGCTACCTCTCTTCTACTTAAACAGACCTTAAATAGTCTTCC 1808
613 eLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPhe 629
|||||
1809 TCCTGTACTCAAGTGGGCGCATGTATTTAAACAAGTTAATGCTTCC 1858
630 ValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGly 646
|||||
1859 ATTAGAAACGATAAGTGCAGTCCAGATCATGTAATTAGACAAACAGTA 1908
646 rGln 647
|||||
1909 CCAA 1912

```

seq_name: /cgn2_6/prodata/2/ina/6B-COMB.seq:us-09-376-781-24

seq_documentation_block:

Sequence 24, Application US/09376781

Patent No. 6261806

GENERAL INFORMATION:

APPLICANT: Banerjee, Papia T.

APPLICANT: Patience, Clive

APPLICANT: Andersson, Goran K.

TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of

Patent No. 6261806

TITLE OF INVENTION: Use

FILE REFERENCE: 61750-267

CURRENT APPLICATION NUMBER: us/09/376, 781

EARLIER FILING DATE: 1999-08-18

EARLIER FILING DATE: 1998-08-18

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 24
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Nucleotide
; OTHER INFORMATION: sequence of the 3' end of the PERV-D env region.
US-09-376-781-24

```

```

alignment_scores:
  Quality: 1675.00      Length: 489
  Ratio: 4.095          Gaps: 8
  Percent Similarity: 83.640  Percent Identity: 66.871

```

```

alignment_block:
US-09-171-553B-10 x US-09-376-781-24

```

Align seg 1/1 to: US-09-376-781-24 from: 1 to: 1493

```

163 ProGlyLysTyrLysMetMetLysLeu...TyrLysAspLysSer...Cy 177
|||||
4 CCAATTAACCAACCGACCTGGAAAACTCATACAGGTACCATTCCTGGGTG 53
177 sSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluTyrGlyL 194
|||||
54 TTTTCCCTCAGACCTCAGATTTATCTTAAATAGTTTCACCGAAAAAAA 103
|||||
194 ySgInGluAsnIleGlnLysTyrPilleAsnGlyMetSerTyrPglyIleVal 210
|||||
104 AACAAAGAAATATCTTAAATGATTAATGATTAATGATGCTCGGGAATATA 153
|||||
211 PheTyrLys...TyrGlyGlyValGlySerThrLeuThrIleArgLe 226
|||||
154 TATTATACAGGTGGGACAGCAACCAACGCTCCATTCTTAACCATCCGACT 203
|||||
226 uArgIleGluThrGlyThrGluProProValAlaMetGlyProAspLysV 243
|||||
204 TAAATA...AGCCAGCTAGAGCCTCCATGCTATAGAGCAATACGG 250
|||||
243 aLeuAlaGluGlnGlyProProAlaLeuGluProProHisAsnLeuPro 259
|||||
251 TCTTAACGGGTCAAGAAACCCCAACCCAGGACCATCCTCGATATA... 297
|||||
260 ValProGlnLeuThrSerLeuArgProAspIleThrGlnProProSer 276
|||||
298 .....ACTTCTAAATTAAGACCCCTAG...TCTAA 326
|||||
276 nSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 293
|||||
327 CAGCAGCACT..... 336
|||||
293 aProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPhe 309
|||||
337 .....AAAACGGGACAAACAACTTTTGTAGTCTCATCCAGGAGCTTT 378
|||||
310 GlnAlaIleAsnSerThrAspProAspAlaThrSerSerGlyTyrPleu 326
|||||
379 CAAGCTCTTAACTCCAGCACTCCAGAGGCTAACCTTCTTGTGGCTTTG 428
|||||
326 sLeuSerSerGlyProProTyrTyrGluGlyMetAlaValGluArgLys 343
|||||
429 CTTAACTTGGGGCCACCTTACTTAAAGAAATGGCTTAAAAAGAAAAAT 478
|||||
343 heAsnValThrLysGlnHisArgAsnGlnCysThrTyrPglySerArgAsn 359
|||||
479 TCAATGTGACAAAAAACAATAGAGCAATGATCATGAGGATCCCAAAAT 528
|||||
360 LysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysAl 376
|||||
529 AAGCTTACCTTACTGAGGTTTGTGAAAGACACCTCTTAAAGAAAGT 578
|||||
376 aProProSerHisGlnHisLeuGlySerTyrSerThrValValGlyGlnAla 393

```

[illegible]

```

1  APPLICANT: Collins, Mary KL
2  APPLICANT: Weiss, Robin A
3  APPLICANT: Takeuchi, Yasuhiro
4  APPLICANT: Cosset, Francois-Loic
5  TITLE OF INVENTION: Expression systems
6  FILE REFERENCE: 09/011,745
7  CURRENT APPLICATION NUMBER: US/09/011,745
8  CURRENT FILING DATE: 1998-06-22
9  EARLIER APPLICATION NUMBER: PCT/GB96/02061
10 EARLIER FILING DATE: 1996-08-23
11 EARLIER APPLICATION NUMBER: GB9517263.1
12 EARLIER FILING DATE: 1995-08-23
13 NUMBER OF SEQ ID NOS: 29
14 SOFTWARE: PatentIn Ver. 2.0
15 SEQ ID NO 7
16 LENGTH: 6312
17 TYPE: DNA
18 ORGANISM: Artificial Sequence
19 FEATURE:
20 OTHER INFORMATION: Description of Artificial Sequence: Portion of
21 FEATURE:
22 OTHER INFORMATION: construct
23 FEATURE:
24 NAME/KEY: misc_feature
25 LOCATION: (4058)
26 OTHER INFORMATION: n is any nucleotide
27 FEATURE:
28 NAME/KEY: misc_feature
29 LOCATION: (4059)
30 OTHER INFORMATION: n is any nucleotide
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: (4060)
34 OTHER INFORMATION: n is any nucleotide
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: (4061)
38 OTHER INFORMATION: n is any nucleotide
39 FEATURE:
40 NAME/KEY: misc_feature
41 LOCATION: (4246)
42 OTHER INFORMATION: n is any nucleotide
43 FEATURE:
44 NAME/KEY: misc_feature
45 LOCATION: (4247)
46 OTHER INFORMATION: n is any nucleotide
47 FEATURE:
48 NAME/KEY: misc_feature
49 LOCATION: (4248)
50 OTHER INFORMATION: n is any nucleotide
51 FEATURE:
52 NAME/KEY: misc_feature
53 LOCATION: (4249)
54 OTHER INFORMATION: n is any nucleotide
55 US-09-011-745-7

```

```

alignment_scores:
    Quality: 1544.50      Length: 660
    Ratio: 3.300          Gaps: 15
    Percent Similarity: 70.909      Percent Identity: 46.970

alignment_block:
    US-09-171-553B-10 x US-09-011-745-7  ..

Align seg 1/1  to: US-09-011-745-7  from: 1  to: 6312

47  ILeaSPSeSerAnSProHISArGPLeuSeSProThrTrpLeuIleI 63
    ::::::::::::::::::::::::::::  :::::
1675 CTCGAAATATAGAACCCCGACGACCATGACCTCCTACTGGCAGTACT 1724
    ::::::::::::::::::::  :::::
63 eaSPProASpThnClYalThrValAsnSerThrArgGlyValAlaPro 80
    ::::::::::::::::::::  :::::
1725 GTCC...CAAACTGGAGACCTGTCTTGATACAAAGCGATCGAGCCC 1771

```

80 rgglyThrTrpProgluLeu.....HispheCysleuArgLeu... 93
1772 CTTGGACTTGGTGCCCACTTAACCTGATGTATGTGCTTGGCGCT 1821
94leuAsnProAlaVally 99
1822 ACTCTTGAAGTCCCTGGGATATCCCGGGAAACGATGTCTGCTCTAAACG 1871
99 sserThrProProasn..... 104
1872 ACTGAGACTCCGAGACTACGACTATACCGCTTATAGCAATCACT 1921
105LeuValArgSer 108
1922 GGGAGCATAGAGGTGACGTAACCTCGGCTAGGACTAGAAATGGCAACG 1971
109 TyrGlyPheTyrCysCysPro.....GlyThrGluLysGlu 121
1972 TCTACCTTCTAGTATGTCTCCGGGATGGCGGACCTTTTCAGAACTAG 2021
121 sTyrCysGlySerGlySerGlyLeuSerPheCysArgArgrTrpSerCysValT 138
2022 AAGTGGCGGGGGCTAGATCCCTATACTGTAAGAATGGGATTTGTGAGA 2071
138 hrSerAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgVal 154
2072 CCACGGGACCGGTTATTGG...CTATTAATCTCTCAAAAGACCTCAT 2118
155 LysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLysLe 171
2119 ACTGTAATAGGAGCAAAATAGCAATGAGACTCAAAATTTTCAACAGTG 2168
171 uTyrLysAspLysSerCysSerProSerAspLeuAspTyrLeuLysIles 188
2169 TCACGAGACCGGCTGTGTATACCC.....CTTAATAATAG 2203
188 erPheThrGluLysGlyLysGlnAsnIleGlnLysTrpIleAsnGly 204
2204 ATTTACAGACAAAGAAATTTATCC.....AAGGACTGATTAACGGGA 2247
205 MetSerTrpGlyLysLeuPheTyrLysTyrGlyLysGlyLysSerTh 221
2248 AAAACCTGGGATTAAGATTCTAT...GTGTGTGACATCCAGCGCTACA 2294
221 rLeuThrIleArgLeuArgIleGluThrGlyThrGluProProAlaIam 238
2295 GTTCACATTCGCTTAATAATCACCC.....AACATGCCAGCTGTGGCAG 2338
238 etGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeuGlu... 253
2339 TAGTCCCTGACCTGCTCTGTGTGACACAGACCTCTAGACGCTCCCTC 2388
254ProProHisAsnLeuProValProGlnLeuTh 264
2389 GCTCTCCACCTCTCTCTTCCCAAGGAGGACGCCACCGCATCTCTC... 2436
264 rSerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyL 281
2437CCCGACTCTTAACCTCCACAGCCCTGG 2461
281 euIleProThrAsnThrProArg.....Asn 289
2462 CGACTAGTGCAACAACTCCACGCTGAGAAAAACAATTGTACCTTAAC 2511
290 SerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleG1 306
2512 ACTCG.....CTCCACACAGCGGACGACTTTTGTATCTTTGCA 2555
306 nGlyAlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerC 323
2556 GGGGGCTTCTTAACCTTAATGCTTACCAACCCAGGGGCCAGTGAGCTT 2605

323 yStrPleuCysLeuSerSerGlyProProTyrTyrGlnGlyMetAlaLys 339
2606 GCTGGCTTTGTTGGCCATGAGCCCCCTTATATAGAAAGCAATAGCCGCA 2655
340 GluArgLysPheAsnValThrLysGlnHisArgAsnIleCysThrTrpG1 356
2656 TCAGAGAGAGGTCCGCTACTCCACCGAC...CTTGACCGGTGCGCTGGG 2702
356 ySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys1 373
2703 GACCAAGAGAAAGCTCACCTCTACGTAGAGTCTGAGGACAGCGGTGTGCA 2752
373 leGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrVal 389
2753 TAGAAAGTGTGCTTACCCATCAGATCTCTGCAATCAGACCCCTATCC 2802
390 TyrGlnGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTr 406
2803 ATCAATTCTCCGAGACATCAGTATCTGCTCCCTCCCAACCATAGCTG 2852
406 pTrPAlaCysAsnThrGlyLeuThrProCysValSerThrSerValPhe 423
2853 GTGGGCTTGACAGCACTGGCTCACCTTGCCTCTCCACCTCAGTTTTR 2902
423 snGlnSerLysAspLeuCysValMetValGlnIleValProArgValTyr 439
2903 ATCAGACTAGAGATTCTGTATCCAGGTCCAGTGATCTCTGCCATCTAT 2952
440 TyrHisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnAr 456
2953 TACTATCTCTGAAGAGTTGTTGTACAGGCTTATGACAATTCTCACCCAG 3002
456 gProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLys 473
3003 GACTAAAGAGAGGCTGTCTACTTACCTCAGCTGTTTACTGGGGTGG 3052
473 1YThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyPro 489
3053 GAATTCACGCGGGGATATAGTACTGTTCAAACTGCTTAATTAAGAACCT 3102
490 GlnGlnLeuGluLysGlyLeuGlyGluLeuHisAlaIleMetThrGluAs 506
3103 ATGACCTCCAGCAAGCGCTGACAGACGCTCCAGATGCGCATAGATCGTA 3152
506 pleuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuThs 523
3153 CTTCCGGGCTCCAAAGCTCAGTCACTCAGCAAGTTAGAGACTCACTGACTT 3202
523 erLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPhe 539
3203 CCTGTCCGAGGTAGTGTCCAAATAGGAGAGGCTTGACTTGTCTGTTT 3252
540 LeuArgGluGlyLysLeuCysAlaAlaLeuLysGluGluCysCysPheTyr 556
3253 CTAAAGAGAGGTGCTCTGTGTGCGGCCCTAAAGAGAGAGTGTGTTT 3302
556 rValAspHisSerGlyAlaAlaIleArgAspSerMetAsnLysLeuArgLys 573
3303 CATAGACACTCAGTGTGACGTACGGGATCTCAGTGAAGAACTCAAGAAA 3352
573 ySLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrpPheGlu 589
3353 AACTGTGATTAAGAGAGTATAGAGCGCCACAAAAGCAAACTGTATGAA 3402
590 GlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuTh 606
3403 GGATGTGTTCAATTAACCTCTGTGTGCTCACTACCTGTCTTCAACATGCG 3452
606 rGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCysLeu 623
3453 TGGGCCCTTAATACTCTCTCTGTGTGCTTCACTCCGGGGCATGATGA 3502
623 leAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 639

[illegible]

```

seq_name: /cgn2.6/ptodata/2/lna/6A_COMB.seq:US-09-075-272-1
seq_documentation_block:
: Sequence 1, Application US/09075272
: Patent No. 6136598
GENERAL INFORMATION:
APPLICANT: MILLER, A. DUSTY
APPLICANT: WOLGAMOT, GREG
APPLICANT: BONHAM, LYNN
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
TITLE OF INVENTION: PACKAGING CELL LINES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
City: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,272
FILING DATE: 08-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,140
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Poor, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: 14538A-003710
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 467-9600
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8655 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-075-272-1

alignment_scores:
Quality: 1478.50 Length: 728
Ratio: 3.080 Gaps: 24
Percent Similarity: 65.934 Percent Identity: 44.368

Alignment_block:
US-09-171-553B-10 x US-09-075-272-1 ..
Align seg 1/1 to: US-09-075-272-1 from: 1 to: 8655

6 SerTTPrATgHIS.leuProThrArGtGlyGluProLYsArgLeuArgI 22
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5711 TCCTGGATTCACGCTTCACACAGCGACAGAGGCCGCCGTCACATGAGTAAG 5760
22 lePro.....leuSer 25
|||
5761 AACCCACGAGACACACTGGGCTGTGGAAGCCACTGATAACCTCTTAAAG 5810
26 pheAlaSerIleAlaITrPheLeuThrIleThrProGlnAla 42
||||| ::||| ::||| ::||| ::|||
5811 TTCGCTGGTGGTCGACGAGACCCGCCCTCCATCACCCCTGAGC..... 5849
42 rSerLYsArgLeuIleAspSerSerAsnProHisArgProLeuSerProt 59

```



```

5850 .....CTAGGACACACACCTCATGCCCCAGTTCACAGT 5886
59 htrpLeuIlelleasproAspThrGlyValThrValAsnSerThrArg 75
5887 CTTGGAGAGTGCTTAAT...GAAAGGAGAGCTGTATGGTACCACCT 5933
76 GlyValAlaProArgGlyThrTrpProGluLeu.....HisPheCys 90
5934 GCAGTCCATCCCTGCTGTGGCTGTGCTGATCTCAGACCTGACATCTG 5983
90 s.....LeuArgLys 93
5984 TAAATTAGCAGCAGATCTCCCAATTGGACCTCCCATCATACTAGTACT 6033
93 euIleAsn...ProAlaValLysSerThrProProAsnLeuVal..... 106
6034 TGAATTAACCCACCTCTCTGACAAAGTGTGTCCCAACGGGTAGAGAC 6083
107 .....ArgSerTyrGly 110
6084 ACTACTGATGTTCCGGGCGAGTCTATGAGCTAATCTTAGGGCTCAGCA 6133
110 yPheTyrCysCysProGlyThr.....GluLysGluLysTyrC 123
6134 GTTTATATGTTGCCCTGGTCAGGTCAGAAAGGAAAGCTGCACAAAGANT 6183
123 ySGLyGLySerGlyLysSerPheCysArgArgTrpSerCysValThrSer 139
6184 GTAGAGGGGCATCAGACTACTTTCGGTAATGACATGTAAACACACA 6233
140 AsnAspGlyAspTrpLys.....TrpProIle..... 148
6234 GGGGAAGCCTACTGGAAGCCTTCCGCTGATGAGCATCAGTCACGTAAG 6283
149 .....SerLeuGlnAspArgValLysPheS 157
6284 ACGTGTAAGTGTATGATAGCCAAACAGAGAAAGAAACCATACACA 6333
157 erPheValAsnSerGly.....ProGlyLys 165
6334 AATACTAGATTTCTGGGTGCTGCTTAAATAATACAGCCCCCAGAGACA 6383
166 TyrLysMetLysLeuTyrLysAspLysSerCysSerProSerAspLe 182
6384 .....TGCAAGGTAAATACTGCAACCC..... 6407
182 uAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnLysIleG 199
6408 .....CTACTCATTAATAATTCAGAGAAAGGAAACAGCTCCCTG. 6449
199 IuLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215
6450 ..AGTTGGCTTAAAGGAATAGGTGGGTGGCGAGTATATATTCACATA 6497
216 GlyGlyAlaGlySerThrLeuThrIleArgLeuArgIleGluThrGlyTh 232
6498 AGAGACCTCGGTTTATCTTACGATTAGACTGACAGTAGA.....GA 6541
232 rGluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyP 249
6542 CTTGGCAGTATCATCATAGGAGCCACAAAGGCTCTTAGCGAAACGGCC 6591
249 rOProAlaLeuGluProProHisAsnLeuPro...ValProGluLeuThr 264
6592 CCCCACTTGCACCGGCTCCCGAGAGTCCAGCCGTGCACCTCAGCACA 6641
265 SerLeuArgProAspIleThrGlnProProSerAsnSerThrThrGlyLe 281
6642 ACTTCACGGGCC...TACACAGTAGAGACCTCATTAGACAAACCTTAGC 6688
281 uIleProThrAsnThrProArgAsnSerProGlyValProValLysThrG 298

```

```

6689 CTCCCCACCA.....CTCTAGATACAG 6711
298 LysIleArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSer 314
6712 AAACCCCTGCGCAGCTAGTTCAGGAGCCCTTTTATGTTAAATAGC 6761
315 ThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyP 331
6762 ACTAATCTTAATATGACTCAATCATGTGGTATAGCTATGCTCTTAAACC 6811
331 oProTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysG 348
6812 CCTTATATGAGAGATTCCTCAGACTAGACTACATTAATATATCTTACG 6861
348 IuHisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThr 364
6862 ATCAT...TCTCAATGCTTTGGGAGAGAAACGAGAACTGACTCTCACA 6908
365 GluValSerGlyLysGlyThrCysIleGlyLysAlaProProSerHisG 381
6909 GCAGTTTCCAGAAATGGGCTTTGTTAGGTGAGGTGCCCGCAGATTAATG 6958
381 nHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsnGlnT 398
6959 GCACCTCTGTAAACAGACGATCAAAATATCCGACCTAACAAAGGTGTCAGT 7008
398 yLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLeuThr 414
7009 ATCTAGTGCCCTCCATGACACAGATARGGCTTGCAATACAGGTCTCAGT 7058
415 ProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysValMe 431
7059 CTTGTATATCTATGCTGTGTTTCAATAAGCTCCAAAGATTTGCTATTTT 7108
431 tValGlnIleValProArgValTyrTyrHisProGluGlnValValLeuA 448
7109 AGTTCAGCTTATTCCTAGACTCCGTAATCATGATGATAGCTCATTTCTAG 7158
448 sPGLuTyrAspTyrArgTyrAsnArgProLysArgGluProValSerLeu 464
7159 ACAATTTGAACATCGGGTC...CCCTGAAAGAGAAACCAATTAATCTTGA 7205
465 ThrLeuAlaValMetLeuGlyLeuGly...ThrAlaValIleGlyGlyTh 480
7206 ACGTTACAGAGCTTTAGATGGAGTGGAGTGGCGCACGTGGGTAGGTTAC 7255
480 rGlyThrAlaIleLeuIleThrGlyProGlnGlnLeuGlyLeuGlyLeuG 497
7256 TGGAAACCGCTGCTTAATCCAGACCCCCCGATACTTTGAG..... 7295
497 LysIleLeuHisAlaIleMetThrGluAspLeuArgAlaLeuLysGluSer 513
7296 ..GAATTACGACAGATGATGATCTGATCTCAGAGCTTAAGAACAACCTP 7343
514 ValSerAsnLeuGluGluSerLeuThrSerLeuSerLysValValLeuG 530
7344 ATAACCAAACTAGAGAAATCTTAATCTCTGTGAGGTAGTACAGCA 7393
530 nAsnArgArgGlyLeuAspLeuPheLeuArgGluGlyLysCysA 547
7394 AAATAGGAGGATTAATGATTTGTTATTTCTTAAGAAAGAGGACTTTGTG 7443
547 IAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIle 563
7444 CTGCTCCAAAGAGAGATGTTTATGATGCGACGACCTGGGAGTGATC 7493
564 ArgAspSerMetAsnLysLeuArgLysLysLeuGluArgArgArgGly 580
7494 AAAGATTCTATGGCCAAACTTAGAGAACGCGCTGATTTAGCTAAAGAGA 7543
580 uArgGluAlaAspGlnGlyTrpPheGluGlyLysTrpPheAsnArgSerPro 597
7544 GAGAGAGAGCAACAGAGATGGTTTGAAGCTGTTAATATAGTCCCTCT 7593

```

[illegible]

6636 CCCCTACCC.....AACCTGCCAAGTCTCCCGCCGCTC 6670
 275 rAsnSerThrThrGlyLeu...ProThrAsnThrProArgAsnSerp 291
 6671 TAGTTCGACTCCACATTGATTCCTCCGCTCCACCTCCACCTAGAGCC 6720
 291 rGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
 6721 CG.....CCAGCAGGAACGGAGACAGATTACTAATCTAGTACAGGA 6764
 308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerGly 324
 6765 GCTTACCAAGGACCTCAACCTTCAACCTGATTAACCTCAAGGCTGCTG 6814
 324 PLeuCysLeuSerSerGlyProProTyrTyrGlnGlyMetAlaLysGlu 341
 6815 GTTATGCTTACTGTCTGACCCCTTATAGAGGGGGTTCCTCTAG 6864
 341 rGlySPheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySer 357
 6865 GTACTTATTCACACCATACCTCTGCCAGCTACTGCTCCGTGGCTCC 6914
 358 ArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrGlyIleG 374
 6915 CAACACAAGCTGACCTGCTGCGAAGTACTGACGGGACTCTGCATAGG 6964
 374 yLysAlaAlaProSerHisGlnHisLysCysTyrSerThrValValTyrG 391
 6965 AACAGTCCCAAAACTCACCAGGCCCTGTGCAACACTACCTT.....A 7008
 391 LglAlaHisSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTr 407
 7009 AGGAGAGGAAAGGCTTACTATCTACTGTGCCCCACAGGACATATGTG 7058
 408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
 7059 GCATGTACACTGACTGACTCTCCTTACCTGCTTGCACCGTCTTAATCG 7108
 424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTyr 441
 7109 CACACGTACTATTTGCGTTCGGAATATGCCCCAGGCTACCTAC 7158
 441 lSPROGluGluValValLeuAspLysPylTyrArgTyrAsnArgPro 457
 7159 ATCTCCCAAGTACGCTATAGCCAGTTTGAA...AAATCCCAATAGACAT 7205
 458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 471
 7206 AAAAGAGAACAGGTGCTTACCTTGCCCTTATTTAGGTGGCTAAC 7255
 472 ...LeuGlyThrAlaValAlaGlyValGlyThrAlaAlaLeuIleT 487
 7256 TATGGGTGGCATGCGCGGGAGTAGGACAGGAACCTACCCCTGGTGC. 7304
 487 hrGlyProGlnGlnLeuGluLysGlyLeuGlyLysIleAlaAlaMet 503
 7305 ..GCCACCCACAGTTTACG.....CAGCTCCATGCTGCCGCTA 7340
 504 ThrGluAspLeuArgAlaIleLysGluSerValSerAsnLeuGluLys 520
 7341 CAAGATGATCTCAAAAGAGTCAAAAGTCAATTAATTAAGCTAGAAAGTC 7390
 520 rLeuThrSerLeuSerGluValValLeuGlnHisArgArgGlyLeuAsp 537
 7391 TCTTACTTCGTTGCTGAGGTGTTAGTCGAATTCGACAGAGCCTAGACC 7440
 537 eulePheLeuArgGluGlyLysCysAlaAlaLeuLysGluGlyCys 553
 7441 TGTGTTCTTAAAGAGGAGGAGTGTCTGCTGCTTAAAGAGAAATGT 7490
 554 CysPheTyrValAspHisSerGlyAlaAlaLeuArgAspSerMetAsnLys 570
 7491 TGTTCCTATGCTGACATACAGGCTTAGTAGATAGTATGCGCAATTT 7540

570 uArgLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGlyT 587
 7541 AAGAGAGAGACTCTCTCAGACAGCAAAACTATTGTAGTGCAGCAGGAT 7590
 587 rPheGluGlyTyrPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
 7591 GGTTCGAGAGATGGTTTAAACAGATCCCTCGTTTACCAAGCTTGAATCC 7640
 604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyP 620
 7641 ACCATATGAGGCGCTCTCATTAATCTCTACTTAATTCCTTTTGGACC 7690
 620 oCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAla 637
 7691 CTCATTCCTTAATCGATTAATGTTCAATTTGTTAAAGACAGATCAGTAG 7740
 637 aGlnIleMetValLeuArgGlnGlnTyrGlnGlyLeu 649
 7741 TCCAGGCTTATGCTCTGACTCAACATATACCAACAGCTA 7778
 seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-886-642-8
 seq_documentation_block:
 ; Sequence 8, Application US/08886642
 ; Patent No. 5952474
 ; GENERAL INFORMATION:
 ; APPLICANT: Pfizer, Abraham
 ; APPLICANT: Kayman, Samuel
 ; TITLE OF INVENTION: FUSION GLYCOPROTEINS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 45 Rockefeller Plaza, Suite 2800
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows95
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/886,642
 ; FILING DATE: 01-JUL-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/110,300
 ; FILING DATE: 20-AUG-1993
 ; APPLICATION NUMBER: 07/938,100
 ; FILING DATE: 28-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hone, William J.
 ; REGISTRATION NUMBER: 26,739
 ; REFERENCE/DOCKET NUMBER: 07763/010002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212/765-5070
 ; TELEFAX: 212/258-2291
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8323 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; US-08-886-642-8
 alignment_scores:
 Quality: 1329.00 Length: 696
 Ratio: 2.973 Gaps: 23
 Percent Similarity: 64.224 Percent Identity: 42.816

```
alignment_block;
```

US-09-171-553B-10 x US-08-886-642-8

Align seg 1/1 to: US-08-886-642-8 from: 1 to: 8323

16 GluProLysArgLeuArgIleProLeuSerPheLeuLeuIleAlaTrp 32
:::||||| ::| |||||::| ::|
5817 GACCCGGGAGCCTCCTAATCCCTTAAATCTC.....TT 5851

32 LeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAsp 49
|||||::|
5852 CCTCTCTCTC.....AAAGGGCCAGATCCGACGACCCG 5886

49 SerAsnProHisArgProLeuSerProThrTrpLeuIleIleAspPro 65
:::|||||::|::| ::| ||||| ::|
5887 GCTCCAGCCCTCACCAAGGCTTACAACTATACCTGGGAAGGACCAATGGG 5936

66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGlyThr 82
|||||::|
5937 GATGGG...GAGACAGTATGGGCAATATCAGGCAACCCCTCTGTGGAC 5983

82 TrpTrpProGlnLeu..... 87
|||||::|
5984 TTGGTGGCCAGTCTCTACCCCAATTTGTGTATGTAGTCTCAGTGGG 6033

88HisPhe..... 89
|||||::|
6034 CGCCCCACTGGGGGCTAGAGTATCAGGCCCTTATCTCGCCCGGGG 6083

90CysLeuArgLe 93
|||||::|
6084 CCCCCCTGTGTCTCAGGAGCAGCGGGAACGTTGACGGCTGTGCCAGAA 6133

93 uIleAsnProAlaValLysSerThrProPro.....A 104
|||
6134 CTGCACAGAGCCCTTACCTCCTCCTACCCCTCGTGCAACACTGCCTGA 6183

104 snLeuValArg.....SerTyrGlyPheTyr 112
|||||::|
6184 ACAGACTTAGCTGAGCAACAGTAACTCATTAATCAAGTACGAGGATTTAT 6233

113 CysCysProGlyThrGlyLys.....GluLysTyrCysGlyLys 126
|||||::|
6234 GTCGTGCCCGGTACATCATGCCCCCGGAAGCAAGTCTGTGGGGGTCC 6283

126 rGlyGluSerPheCysArgArgTyrSerCysValThrSerAsnAspGly 143
:::|||||
6284 AGACTCTCTTACTGTGCTCTTGGGGGTGACAGAACCCGGTAAAGTAT 6333

143 sPTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
|||||::|
6334 ACTGGAG...CCCTCTCTTCTTGGGACTACATCA.....GTA 6371

160 AsnSerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAsp...Ly 175
:::|||||
6372 GACAAACAATCTCACTTAACAGGCTGTTCAGTATGCAAGACAAATAA 6421

175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGly 192
|||||::|
6422 GTGGTCAATCCC.....TTGGCTATCCGGTTTACAAACG 6456

192 ySgLYLysGlnGlnAsnIleGlnLysTrpIleAsnGlyMetSerTrpGly 208
|||||::|
6457 CCGGAAACAG.....GTCACCTCATGGACAGACCCAGACATATGGGGT 6500

209 IleValPheTyrLysTyrLysGlyLysAlaGlySerThrLeuThrIleAr 225
:::|||||
6501 CTACGCTTTATGTCTGTGACAGACACCCAGGCTTACTTTCGGATCCG 6550

225 GluArgIleGlyThrGlyThrGluProProValAlaMetGlyProAsp 242
:::|||||
6551 ACTCACTTATCAA...AATCTAGACCTCGGATCCCAATAGACCAACCG 6597

275 rAsnSerThrThrGlyLeuIle...ProThrAspThrProArgAsnSer 291
:::|||||
6671 TAGTTCAGCTCCCAATGATTTTCCCGCTCCCACTCCCACTACGCC 6720

291 rGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGly 307
|||
6721 CG.....CCAGCAGAGACGGAGACAGATTAATAATCTGTACAGGA 6764

308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTr 324
|||||::|
6765 GCTTACCAAGGACTCAACCTTATACCAACCTGATTAACCTCAAGAGTCTG 6814

324 PLeuCysLeuSerSerGlyProProTyrTyrGlyGlyMetAlaLysGlu 341
|||||::|
6815 GTTATGCCCTAGTGTCTGAGACCCCTATTAAGAGGGGTTCCCTCTAG 6864

341 rGlySPheAsnValThrLysGlnHisArgAsnGlnCysThrTrpGlySer 357
:::|||||
6865 GTACTTATTCACACATACCTCTGCCACGCTACTGCTCGTGCCCTCC 6914

358 ArgAsnLysLeuThrLeuThrGlyGlnValSerGlyLysGlyThrCysIleG 374
:::|||||
6915 CAACACAGCTGACCTGTCCGAAAGTACATGAGAGGGGACTGTGCATAGG 6964

374 yLysAlaProProSerHisGlnHisLysCysTyrSerThrValValTyrG 391
|||
6965 AACAGTCCCAAAACTCACAGGCGCTGTGCACATCACCTT.....A 7008

391 LuGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnAlaGlyTrp 407
:::|||||
7009 AGGAGAGAAAGGCTTACTATATGTTACCTCCCAACAGACTATGTGG 7058

408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
|||||::|
7059 GCATGTAACTGACATCATCTCCATGCTTATTCGCCACCGTCTTAATG 7108

424 nSerLysAspLeuCysValMetValGlnIleValProArgValTyrTrp 441
:::|||||
7109 CACCACTGACTATGCGTTCGCGGAATATGAGCCAGGCTCACTAC 7158

441 sProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArgPro 457
|||
7159 ATCTCCAGTACGTCTATATGACAGTTTGA...AAATCCCATAGACAT 7205

458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly..... 471
|||||::|
7206 AAAAGAAACACAGTCTTAACCTTGCGCTTATTTATAGTGGGCTAAC 7255

472LeuGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIle 487
|||
7256 TATGGGTGCATGCGCCGCGGAGTATGAGACAGAACTACCGCCCTGTC 7304

487 hTrGlyProGlnGlnLeuGlyGlyLeuGlyLeuHisAlaIleMet 503
:::|||||
7305...GCCACCCAGCAGTTTCAG.....CAGTCCATGCTGCCGTA 7340

504 ThrGluAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnGlu 520
:::|||||
7341 CAAGATGATCTCAAAGAGTGCAGAAAGTCAATTACTAACCTGAAAGAGC 7390

520 rLeuThrSerLeuSerGlyValValLeuGlnAsnArgArgGlyLeuAsp 537
|||||::|
7391 TCTTACTGTGTGTCGAGGTGTACTGACAGATGACAGAGGCTTAGAC 7440

```

|||||
7441 TGTGTTCTTAAAGAGGAGGAGCTGTGTCTGCCCTTAAAGAGAGATGT 7490
554 CysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLe 570
|||||
7491 TGTTCATCTGCTGACCATACAGGCGCTAGTAAGATAGTATGGCCAAAT 7540
570 uArgLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGly 587
|||||
7541 AAGAGAGAGACTCTCTCAGAGACAAAACATATTGAGTCGAGCCAAAGAT 7590
587 rPheGlnGlyTyrPheAsnArgSerProThrMetThrThrLeuSer 603
|||||
7591 GGTTCGAGGAGATGTTTAAACAGATCCCGCTGTTACACAGTGTATATCC 7640
604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPr 620
|||||
7641 ACCATCATAGGGCGCTCTCATATACCTCTACTAATCTCGCTTTTGACC 7690
620 cCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAla 637
|||||
7691 CTGCATCTTATATCATAGTATGTAATGTTAAAGACAGATCTCAGTAG 7740
637 aGlnIleMetValLeuArgGlnGlnGlnGlnGlyLeu 649
|||||
7741 TCCAGGCTTTAGTCTGACTCAACATATCCACAGCTA 7778

```

seq_name: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:PCT-US93-08041-8

```

seq documentation block:
Sequence 8, Application PC/TUS9308041
GENERAL INFORMATION:
APPLICANT: The Public Health Research Institute of the City of
APPLICANT: New York, Inc.
TITLE OF INVENTION: FUSION GLYCOPROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Davis Hoxie Faithfull and Hapgood
STREET: 45 Rockefeller Pl.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08041
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,100
FILING DATE: 28-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 20-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jacobs, Seth H.
REGISTRATION NUMBER: 32,140
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 8323 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

```

: ANTI-SENSE: NO
PCT-US93-08041-8
alignment_scores:
Quality: 1329.00 Length: 696
Ratio: 2.973 Gaps: 23
Percent Similarity: 64.224 Percent Identity: 42.816
alignment_block:
US-09-171-553B-10 x PCT-US93-08041-8
Align seg 1/1 to: PCT-US93-08041-8 from: 1 to: 8323
16 GluProLysArgLeuArgGlyLeuProLeuSerPheIleAlaThrPhe 32
|||||
5817 GACCCGCGGAGACCTCTAATCCCTTAATTCCTC.....TT 5851
32 eLeuThrLeuThrIleThrProGlnAlaSerSerLysArgLeuIleAsp 49
|||||
5852 CCTGTCTCTC.....AAGGGCCAGATCCGACGACCCG 5886
49 eSerAsnProHisArgProLeuSerProThrTyrLeuIleAspPro 65
|||||
5887 GCTCCAGCCCTCACAGGTCTACACATTAATCTGGAGATGACCAATGGG 5936
66 AspThrGlyValThrValAsnSerThrArgGlyValAlaProArgGly 82
|||||
5937 GATCGG...GAGACAGTATGGCAATATCAGGACACACCCTCTGTGGAC 5983
82 rTyrTrpProGluLeu.....CysLeuArg 87
|||||
5984 TTGTGTGCCCGCTCTCACCACCCAGATTTGTGTATGTAGCTTCAGTGGGC 6033
88 .....HisPhe.....CysLeuArg 89
|||||
6034 CGCCCACTGGGGGCTAGAGTATCAGCCGCCATTCCTCGCCCGGGGG 6083
90 .....HisPhe.....CysLeuArg 93
|||||
6084 CCCCCTTTGTTGTCACAGGAGACAGCGGACGTTGACGCTGTGCCAGACA 6133
93 uIleAsnProAlaValLysSerThrPro.....A 104
|||||
6134 CTGCACAGAGCCCTTGACCTCCTCAACCCCTCGTGCACACACTGCTGGA 6183
104 snLeuValArg.....SerTyrGlyPheTyr 112
|||||
6184 ACAGACTTAAGCTGACACAGTAACATCAATAATCAAGTGAGGATTTTAT 6233
113 CysCysProGlyThrGluLys.....GluLysTyrCysGlyLys 126
|||||
6234 GTCTGCCCGGGTGCATCGCGCCCGGAGAACCAAGTCTGTGGGGGTCC 6283
126 rGlyGluSerPheCysArgArgTyrPheCysValThrSerAsnAspGly 143
|||||
6284 AGACTCTTCTACGTGCTCTTGGGGCTGCGAGACACCGGATGAT 6333
143 sprLysTyrProIleSerLeuGlnAspArgValLysPheSerPheVal 159
|||||
6334 ACTGGGAG...CCCTCCTCTTCTTGGGACATCAATCA.....GTA 6371
160 AsnSerGlyProGlyLysTyrLysMetLysLeuTyrLysAsp...Lys 175
|||||
6372 GACACACATCTCACCTTAACCGAGGCTTTCAGATGACAAACATATA 6421
175 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGlu 192
|||||
6422 GTGTGCAATGCC.....TTGGCTATCCGGTTTAAACAACG 6456
192 ysgLysLysGlnAlaAsnIleGlnLysTyrPheAsnGlyMetSerTyr 208
|||||
6457 CCGGGAACAAG.....GTCACCTCATGAGACACTGACACTATGTGGGT 6500

```

209 ILeValPheTyrluThrglyGlyAlaGlySerThrLeuThrIleAr 225
 225 gLeuArGlyIleGluThrGlyThrIleProProValAlaMetGlyProAspL 242
 6551 ACTCACTTATCAAA...AACTAGAGACTGGATCCCAATAGGACCAAAAC 6597
 242 ySValLeuAlaGluGlyProProAlaLeuGluProProHISAsnLeu 258
 6598 CCGTCCTGGAGACCAA...CTTTCGTTCCCGCTACCTAAT 6635
 259 ProValProGluLeuThrSerLeuArgProAspIleThrGlnProProse 275
 6636 CCGCTACCC...AAACCTGCCAAGTCTCCCGCCGCTC 6670
 275 rAsnSerThrThrGlyLeuIle...ProHISAsnThrProArgAsnSerP 291
 6671 TACTTGACTCCCACTGATATTTCCCGCCCACTCCCACTCCCACTCC 6720
 291 roGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGly 307
 6721 CG...CCAGAGGAGAGGAGAGAGATTAATCTAGTACAGGGA 6764
 308 AlaPheGlnAlaIleAsnSerThrAspProAspAlaThrSerCysTr 324
 6765 GCTTACCAGGACTCAACCTGACCAACCTGATAAACTCAAGAGTGTG 6814
 324 pLeuCysLeuSerSerGlyProProGlyThrGlyGlnMetAlaLysGln 341
 6815 GTTATGCTTAGTGTGAGACCCCTATTAAGAGAGGAGGTTGCGCTAG 6864
 341 rGlyPheAsnValThrLysGlnHISArgAsnGlnCysThrTrpLysSer 357
 6865 GATACCTATTCACACCATACCTGTCGCCCACTACCTGCTCCGCTCC 6914
 358 ArgAsnLysLeuThrLeuThrGlnValSerGlyLysGlyThrCysIle 374
 6915 CAACACAAAGCTGACCTGTCGAAAGTGAAGTGAAGGAGGAGTGTGATG 6964
 374 yLysAlaProProSerHisGlnHISLeuCysTrpSerThrValAlaTrp 391
 6965 AACAGTCCCAAAACTCACAGGCTGTGCAACCTACCT...A 7008
 391 IuGlnAlaSerGlnAsnGlnTrpLeuValProGlyTrpAsnArgTrp 407
 7009 AGGCAAGGCAAGGCTTACTTACTTACTTACTTACTTACTTACTTACT 7058
 408 AlaCysAsnThrGlyLeuThrProCysValSerThrSerValPheAsnG 424
 7059 GCATGTACACTGAGTCACTCATGCTATGCTACCGGCTGCTAATCG 7108
 424 nSerLysAspLeuCysValMetValGlnIleValProArgValTrpH 441
 7109 CAACCTGACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7158
 441 ISProGluGluValValLeuAspGluTrpAspTrpArgTrpAsnArgPro 457
 7159 ATCTCCCAAGTACGTCTATAGCCAGTTGAA...AAATCCCATAGACAT 7205
 458 LysArgGluProValSerLeuThrLeuAlaValMetLeuGly... 471
 7206 AAAAGAGAACAGTGTCTTAACCTTATATATAGTGGGCTAAC 7255
 472 ...LeuGlyThrAlaValAlaGlyValGlyThrGlyThrAlaLeuIle 487
 7256 TATGGGTGGCATGCCGCGGAGTAGGAGGAGCACTACCCGCTGCTC 7304
 487 ThrGlyProGlnGlnLeuGluLysGlyLeuGlyGlnLeuHISAlaLam 503
 7305 ...GCCACCCAGAGTTTCAG...CAAGTCCATGCTGCTGCTA 7340

504 ThrGluAspLeuArgAlaLeuLysGluSerValSerAsnLeuGluLys 520
 7341 CAAGATGATCTCAAGAAAGTCAAAAGTCAATTAATTAATTAATTAAT 7390
 520 rLeuThrSerLeuSerGluValValLeuGlnHISAsnArgGlyLeuAspL 537
 7391 TCTTACTTCTGTTGCTGAGGTGTGACTGCAAGAACGACGAGGCTTAG 7440
 537 eulePheLeuArgGluGlyLeuCysAlaAlaLeuLysGluLys 553
 7441 TGTTCCTTCAAAAGAGGAGGAGTGTGTGCTGCTGCTGCTGCTGCTG 7490
 554 CysPheTyrluAspHisSerGlyAlaIleArgAspSerMetAsnLys 570
 7491 TGTTCCTTCAAAAGAGGAGGAGTGTGTGCTGCTGCTGCTGCTGCTG 7540
 570 uArgLysLysLeuGluArgArgArgGluArgGluAlaAspGlnGly 587
 7541 AAGAGAGAGTCTCTCAGAGACAAAACTATTGAGTCGAGCCAGGAT 7590
 587 rPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSer 603
 7591 GTTTCGAGAGTGTGTTACAGATCCCTGTTGTTACAGTGTGATATCC 7640
 604 AlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrVal 620
 7641 ACCATCATGAGGCGCTCATATATATCTACTAATTTGCTTTTGAGCC 7690
 620 cCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAla 637
 7691 CTGCATTCCTTATGATGATTGTTCAATTTGTTAAAGACAGAGTCTCAG 7740
 637 aGlnIleMetValLeuArgGlnGlnTrpGlnGlyLeu 649
 7741 TCCAGGCTTATGCTCTGACTCAACATTAACACGAGCTA 7778

seq_name: /cgn2_6/prodata/2/lna/6B_COMB.seq:us-09-011-745-6

seq_documentation_block:

? Sequence 6: Application US/09011745
 ? Patent No. 6165715
 ? GENERAL INFORMATION:
 ? APPLICANT: Collins, Mary KL
 ? APPLICANT: Weiss, Robin A
 ? APPLICANT: Takeuchi, Yasuhiro
 ? APPLICANT: Cosset, Francois-Loic
 ? TITLE OF INVENTION: Expression systems
 ? FILE REFERENCE: 09/011,745
 ? CURRENT APPLICATION NUMBER: US/09/011,745
 ? EARLIER FILING DATE: 1998-06-22
 ? EARLIER APPLICATION NUMBER: PCT/GB96/02061
 ? EARLIER FILING DATE: 1996-08-23
 ? EARLIER APPLICATION NUMBER: GB9517263.1
 ? NUMBER OF SEQ ID NOS: 29
 ? SOFTWARE: Patent In Ver. 2.0
 ? SEQ ID NO 6
 ? LENGTH: 6061
 ? TYPE: DNA
 ? ORGANISM: Artificial Sequence
 ? FEATURE:
 ? OTHER INFORMATION: Description of Artificial Sequence: Portion of
 ? FEATURE:
 ? NAME/KEY: misc_feature
 ? LOCATION: (3807)
 ? OTHER INFORMATION: n is any nucleotide
 ? FEATURE:
 ? NAME/KEY: misc_feature
 ? LOCATION: (3808)
 ? OTHER INFORMATION: n is any nucleotide
 ? FEATURE:
 ? NAME/KEY: misc_feature

```

? LOCATION: (3809)
? OTHER INFORMATION: n is any nucleotide
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3810)
? OTHER INFORMATION: n is any nucleotide
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3995)
? OTHER INFORMATION: n is any nucleotide
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3996)
? OTHER INFORMATION: n is any nucleotide
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3997)
? OTHER INFORMATION: n is any nucleotide
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3998)
? OTHER INFORMATION: n is any nucleotide
?
US-09-011-745-6

```

Quality:	1284.00	Length:	734
Ratio:	2.828	Gaps:	26
Percent Similarity:	61.853	Percent Identity:	41.008

```
alignment_block:
nc-08-171-553p-10 v nc-08-011-745-6
```

Align seg 1/1 to: US-09-011-745-6 from: 1 to: 6061

```

2 HisProthLeuSerThrArgHisLeuProThrArgGlyGlyGlyProly 18
1186 CATGCAGAGCTTGGATACACGCCCC...CCAGCTGAAGGCTCCGACCCCG 1332
18 sarGLeuArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrL 35
1333 G.....GGGAGGACCAATCCCT 1349
35 euThrIleThrProGlnAlaSer...SerLysArgLeuIleAspSerSer 50
1350 AGACTGTACATGGCGGGCTTCACAGCTCTCAAAACCCCTTAAAAATAAGGTT 1399
51 AsnProHisArgProLeuSerPro..... 58
1400 AACCCGCGAGGCCCCCTAATCCCTTAATTCCTTGATGCTCAGAGGGGT 1449
59 .....Thr 60
1450 CAGTACTGCTGGCCGGCGGCTCCAGTCACGTCATCAATCATATATATACACT 1499
60 rPLeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGly 76
1500 GGGAGGTACCAATGAGATCGG...GAGACGGTATGGGCAACTCTTGGC 1546
77 ValAlaProArgGlyThrTrpTrpProGluLeu..... 87
1547 AACCAACCTCTGTGACGCTGGGGCTGACCTTACCCAGATTATGAT 1596
88 .....HisPheLysLeuArgLeuIleAsnPro 97
1597 GTTACGCCACATGAGCACAATTATTATGGGGGTAAATAATCATCCCTT 1646
97 laValLysSerThrProPro..... 103
1647 TT.....TCTTCTCCCGGGGGCCCTGTGTGCTCAGGGGCGACAGC 1690
103 ..... 103

```

1691 CAGAGCGTGTTCACAGAGACTGCGAACAACCTTAACTTCCTCCACCCCTCG 1744

104AsnLeuValArg..... 107

1741 GTGCACACTGCCTGGACACAGCATCAACCTAGACCAGACAATCATAAAT 1790

108 ..SerTyrGlyPheTyrCysCysProGlyThrGluIys.....Glu 120

1791 CAATGAGGATTTTATCTTTTCCCGGGGCCACAGCCCGCCGAGATTC 1840

121 LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTyrSerCysVa 137

1841 AAGTCATGTGGGGGTCAACACCTCTTCACTGCTGCTATGTTGGGCTGTGA 1890

137 LTrpSerAsnAspGlyAspTrrPlySerProIleSerLeuInsPrry 154

1891 GACACCCGGTAGAGCTTACTGTAAG...CCCTCCATCATGAGAT.... 1933

154 aLysPheSerPheValAsnSerGlyProGlyLysTyrLysMetMetLys 170

1934TTTCATCACAGTAAACAACATCTACCTCGACAGCGTGTCCAG 1978

171 LeuTyrLysAsp...LysSerCysSerProSerAspLeuAspTyrLeu 186

1979 GTATGCAAGATTAATTAAGTGGTGCACACCC.....TTAGT 2013

186 sLleSerPheThrGluLysGlyLysGlnGluAsnLleInsTrrPleA 203

2014 TATTCGGTTTACGACGCCGGGAGACGG.....GTTACTTCCTGACCA 2057

203 snGlyMetSerTrrPglyLysValPheTyrLysTyrGlyGlyValGly 219

2058 CAGAGCATTTACTGGGCTTACGTTGTGTATGTCTCCGCAAGATCCAGG 2107

220 SerTLeuThrLleArgLeuArgLleThrGlyThrGluProProVa 236

2108 CTTACATTTGGATCCGACCTCAGATACCAA...AATCTAGAGACCCCGGT 2154

236 lAlaMetGlyProAspLysValLeuAlaGluInsGlnIlyProProAlaGns 253

2155 CCACAAATGAGGCCAAACCCGTTCTGGCAACCAACAGCA.....CTCT 2198

253 lIuProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAsp 269

2199 CCAAGCCCAA.....CCTGTTAATCG.....CCTTCA 2227

270 lLleThrGlnProProSerAsnSerThrThrGlyLeuLleProThrAsnTh 286

2228 GTCCACCAACACCCACAGTGGG.....ACTCCTCTCTCCCTACCCAACT 2271

286 rProArgAsnSerProGlyValProValLysThrGlyGlnArgLeuPheS 303

2272 TCCA.....CCGGCGGGAGGAAATAGCGTGGCTTAA 2303

303 eLeuLleGlnGlyAlaPheGlnAlaIleAsnSerThrAspProAspIa 319

2304 ACTTAGTAGACGAGGCCCTACCAAGCCCTTAACCTACACAGTCTGACAAA 2353

320 ThrSerSerTrrPleuLysLeuSerSerGlyProProTyrTyrGluG 336

2354 ACCCAAGAGTCTGGTTGTGTCTAGTAGGGGAGACCCCTTACTAGAAAG 2403

336 yMeTAlaLysGluArgLysPheAsnValThrLysGlnHisArgAsnGlnC 353

2404 GGTTCGCGCTCGGTGGTAGACCTACACCAACATACCTGTGCTCCAGCCAACT 2453

353 ystHtrrPglySerArgAsnLysLleThrLeuThrGlyValSerGlyLys 369

2454 GCTCCCGTGGGCTCCCAACAACTTACCCCTGCCAAGTGCACGGACAG 2503

370 gLysThrCysLleGlyLysAlaProProSerHisGlnHisLysCysTyrSe 386

2504 GGACTGTCTGCTTAGGAGCAGTTCCTCCAAACACATCAGCCGCTTATGTAAATAC 2555

```
386 rThrvAlValTyrGlInAlaSerGluAsnGlnTyrLeuValProGlyT 403
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2554 CACCAGACAGACAGACAGGAGGTC...TATTATCTAGTGGCCCTA 2597
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
403 yRAsnArGTrPrAlaCysAsnThrGlyLeuThrProCysValSerThr 419
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2598 CAGGTACCATGTGGCTGTGTAGTACGGGCTTACTCCATGCATGCCACC 2647
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
420 SerValPheAsnGlnSerLysAspLeuCysValMetValGlnIleValPr 436
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2648 ACCATACAGAACCTTACCACATGATTTGTCTTGTGAACTCTGGCC 2697
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
436 oArValTyrTyrHisProGlnIleValValLeuAspGlnTyrAspTyrA 453
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2698 AAGAGTCACCTATTCATTCCTCCCGCTATGTTACGGCTGTTTGAG...A 2744
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
453 rGTYrAsnArGPrProLysArgIuProValSerLeuThrLeuAlaValMet 469
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2745 GATCCACCGACACAAAGAGACGGGTGTGTTACCTGGCCCTATTA 2794
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
470 LeuGly.....LeuGlyThrAlaValGlyValGlyThrGlyTh 482
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2795 TTGGGTGGACTAACCATGGGGGGAATGCGCTGGAATAGGAACAGGGAC 2844
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
482 rAlaAlaLeuIleThrGlyProGlnGlnLeuGlnLysGlyLeuGlyLul 499
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2845 TACTGCTTAATG...GCCACTCAGCAATTCAG.....CAGC 2879
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
499 euHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValSer 515
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2880 TCCAGGCCCGACTACAGATGATCTCAGGAGAGTTGAAAAATCATCTCT 2929
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
516 AsnLeuGlnGluSerLeuThrSerLeuSerGluValValLeuGlnAsnAr 532
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2930 AACCTAGAAAGCTCTCTCCTCCTGCTGAAAGTTGCTTACAGAAATCG 2979
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
532 gArGlyLeuAspLeuPheLeuArgGlnGlyLeuGlnLysAlaAlaL 549
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2980 AAGGGGCTTAGACTGTATTCTTAAAGAAAGAGGGCTGTGTGCTGCTC 3029
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
549 euLysGlnGluCysCysPheTyrValAspHisSerGlyAlaIleArgAsp 565
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3030 TAAAGAAAGATGTTGCTCTTATGGGAGCCACACAGACTAGTAGAGAC 3079
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
566 SerMetAsnLysLeuArgLysLysLeuGluArgArgArgGluArgGln 582
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3080 ACATGGGCCAAATGAGAGAGAGCTTAATCAGAGACAAAGACTGTTGA 3129
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
582 uAlaAspGlnGlyTyrPheGlnGlyTyrPheAsnArgSerProTyrMetI 599
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3130 GTCAACTCAAGATGTTTGTAGGACTGTTTAAACAGATCCCTTGTTTGA 3179
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
599 hrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeu 615
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3180 CCACCTGTATCTACCATTTATGGAGCCCTCATTTGCTACTTATGAT 3229
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
616 LeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGln 632
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3230 TTGCTCTTGGAGCCCTGCATCTTATGATGATTGTTCAATTGTTAAGA 3279
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
632 uArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGlnGlyL 649
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3280 CAGGATCTCAGTATCCAGGCTTTAGTCTCTGACTCAACATTCACACAGC 3329
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
649 eu 649
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3330 TA 3331
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 03:01:22 ; Search time 212.72 Seconds
(without alignments)
21.294 Million cell updates/sec

Title: US-09-171-553B-13

Perfect score: 20

Sequence: 1 gatgctctcctgaccttly 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_MA: *
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	20	100.0	7333	US-08-766-528-2
2	18.4	92.0	3482	US-09-111-085-3
3	18.4	92.0	8060	US-08-766-528-1
4	18.4	92.0	8132	US-08-766-528-3
5	16.8	84.0	72928	US-09-009-913-1
6	16	80.0	23	US-08-766-528-14
7	16	80.0	23	US-08-766-528-15
8	16	80.0	2033	PCT-US93-12464-1
9	16	80.0	2400	US-08-800-2918-7
10	15.4	77.0	4060	US-08-308-949A-1
11	15.4	77.0	31571	US-08-323-443B-1
12	15.4	77.0	53526	US-08-658-136-2
13	15.4	77.0	53577	US-08-658-136-1
14	15.2	76.0	3103	US-08-826-246-3
15	15.2	76.0	3103	US-08-944-495-3
16	15.2	76.0	3103	US-09-126-640-2
17	15.2	76.0	3103	US-08-925-588-3
18	15.2	76.0	3111	US-09-487-444-3
19	15.2	76.0	8655	US-09-075-272-1
20	15.2	76.0	15101	US-08-799-464A-14
21	15.2	76.0	15101	PCT-US95-09927-14
22	15.2	76.0	15108	US-08-157-005-1
23	15.2	76.0	15108	US-08-747-863-1
24	15	75.0	68	PCT-US95-17026-8
25	15	75.0	70	PCT-US95-17026-9
26	15	75.0	3876	PCT-US95-17026-1
27	14.8	74.0	489	US-08-081-072-9

C 28	14.8	74.0	489	1	US-08-449-093A-9	Sequence 9, Appl
C 29	14.8	74.0	569	2	US-08-483-695-44	Sequence 44, Appl
C 30	14.8	74.0	569	2	US-07-965-285-44	Sequence 44, Appl
C 31	14.8	74.0	569	2	US-08-487-231-44	Sequence 44, Appl
C 32	14.8	74.0	569	4	US-09-201-912-44	Sequence 2, Appl
C 33	14.8	74.0	7886	2	US-08-751-189-2	Sequence 2, Appl
C 34	14.8	74.0	7886	4	US-09-060-836-2	Sequence 2, Appl
C 35	14.8	74.0	7886	4	US-09-184-445-2	Sequence 2, Appl
C 36	14.4	72.0	1539	2	US-08-828-596-1	Sequence 1, Appl
C 37	14.4	72.0	1539	4	US-09-360-197-5	Sequence 5, Appl
C 38	14.4	72.0	2567	3	US-08-993-260-4	Sequence 4, Appl
C 39	14.2	71.0	1160	2	US-08-955-138-5	Sequence 5, Appl
C 40	14.2	71.0	1470	6	5187079-3	Patent No. 5187079
C 41	14.2	71.0	3083	1	US-08-346-849-1	Sequence 1, Appl
C 42	14.2	71.0	3083	2	US-08-293-284A-1	Sequence 1, Appl
C 43	14.2	71.0	4020	4	US-09-050-159-130	Sequence 130, App
C 44	14.2	71.0	4024	4	US-09-162-484-18	Sequence 18, Appl
C 45	14.2	71.0	4928	1	US-08-399-561-1	Sequence 1, Appl

ALIGNMENTS

```

RESULT 1
US-08-766-528-2
; Sequence 2, Application US/08766528
; Patent No. 6190861
;
GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF INVENTIONS: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-08-766-528-2
;
Query Match 100.0%; Score 20; DB 4; Length 7333;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 gatgctctcctgaccttly 20
|||||

```

Db 4442 GATGCTCTCTGCTTGG 4461

RESULT 2

US-09-111-085-3
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 3; Length 3482;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgccccttg 20

Db 528 gatgctctctgccccttg 547

RESULT 3

US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8060 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 4; Length 8060;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgccccttg 20

Db 7763 GATGCTCTCTGCTTGG 7782

RESULT 4

US-08-766-528-3
; Sequence 3, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8132 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 4; Length 8132;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgctctctgccccttg 20

Db 5237 GATGCTCTCTGCTTGG 5256

RESULT 5

US-09-009-913-1/C
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:

```

; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-009-913-1

Query Match      84.0%; Score 16.8; DB 3; Length 72928;
Best Local Similarity 90.0%; Pred. No. 32;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gatgctctcctgccttgg 20
    ||||| | ||||| |
Db 46404 GATGCTCTCCATGCCCTTGG 46385

RESULT 6
US-08-766-528-14/c
; Sequence 14, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-14

Query Match      80.0%; Score 16; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gctctcctgccttgg 20
    ||||| | ||||| |
Db 23 GCTCTCCTGCCCTTGG 8

RESULT 7
US-08-766-528-15
; Sequence 15, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/572,645
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-038CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-766-528-15

```

Query Match 80.0%; Score 16; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 gctctcgtcccttg 20
|||||
Db 1 GCTCTCTGCTTGG 16

RESULT 8
PCT-US93-12464-1/c
Sequence 1, Application PC/TUS9312464
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California, et al.
TITLE OF INVENTION: METHOD FOR INHIBITION OF CELL
ADHESION TO RECEPTORS CONTAINING SELECTINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: US
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/12464
FILING DATE: 21-DEC-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOWELLS, STACY L.
REGISTRATION NUMBER: 34,842
REFERENCE/DOCKET NUMBER: PD2929
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: ALPHA-L-FUCOSIDASE
FEATURE:
NAME/KEY: CDS
LOCATION: 19..1401
PCT-US93-12464-1

Query Match 80.0%; Score 16; DB 5; Length 2035;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 gctctcgtcccttg 20
|||||
Db 1851 GCTCTCTGCTTGG 1836

RESULT 9
US-08-800-291B-7
Sequence 7, Application US/08800291B
Patent No. 6153740
GENERAL INFORMATION:
APPLICANT: J.D. Young & C.E. Cass
TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,291B
FILING DATE: 13-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/499,314
FILING DATE: 7-JULY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hallie, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07254/044M01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5099
TELEFAX: 619/678-5070
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2400 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2401
US-08-800-291B-7

Query Match 80.0%; Score 16; DB 3; Length 2400;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 gctctcgtcccttg 20
|||||
Db 2256 GCTCTCTGCTTGG 2271

RESULT 10
US-08-308-949A-1
Sequence 1, Application US/08308949A
Patent No. 5580703
GENERAL INFORMATION:
APPLICANT: Kolin, Robert M.
APPLICANT: Berns, Kenneth I.
APPLICANT: Linden, Ralph M.
TITLE OF INVENTION: Human Adeno-Associated Virus Integration
Site DNA and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,949A
FILING DATE: September 20, 1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,127
FILING DATE: September 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC92-10F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1

Query Match 77.0%; Score 15.4; DB 1; Length 4060;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 tggctctctgcccctt 19
|||||
DB 2970 tggctctctgcccctt 2986

RESULT 11
US-08-323-443B-1/c
Sequence 1, Application US/08323443B
Patent No. 5654170
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURR, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM R.
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,443B
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0372/0A462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 31571 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: PKD1 GENOMIC
US-08-323-443B-1

Query Match 77.0%; Score 15.4; DB 1; Length 31571;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatgctctctgcccct 17
|||||
DB 13938 GCTGGCTCTCTGCCCCCT 13922

RESULT 12
US-08-658-136-2/c
Sequence 2, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURR, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-5415
TELEFAX: 508-872-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match 77.0%; Score 15.4; DB 3; Length 53526;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gatgctctctgcccct 17
|||||
DB 14923 GCTGGCTCTCTGCCCCCT 14907

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,246
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-078-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3103 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 288...1565
OTHER INFORMATION:
US-08-826-246-3

OY      1 gatgctctcctgaccttgy 20
         | | | | | | | | | | | | | |
Db      2031 GGTGTCCTCTCTGCCCTTG 2050

Query Match      76.0%: Score 15.2; DB 3; Length 3103;
Best Local Similarity 85.0%: Freq. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

RESULT 15
US-08-944-495-3
Sequence 3, Application US/08944495
Patent No. 6087477
GENERAL INFORMATION:
APPLICANT: Fald, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,495
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

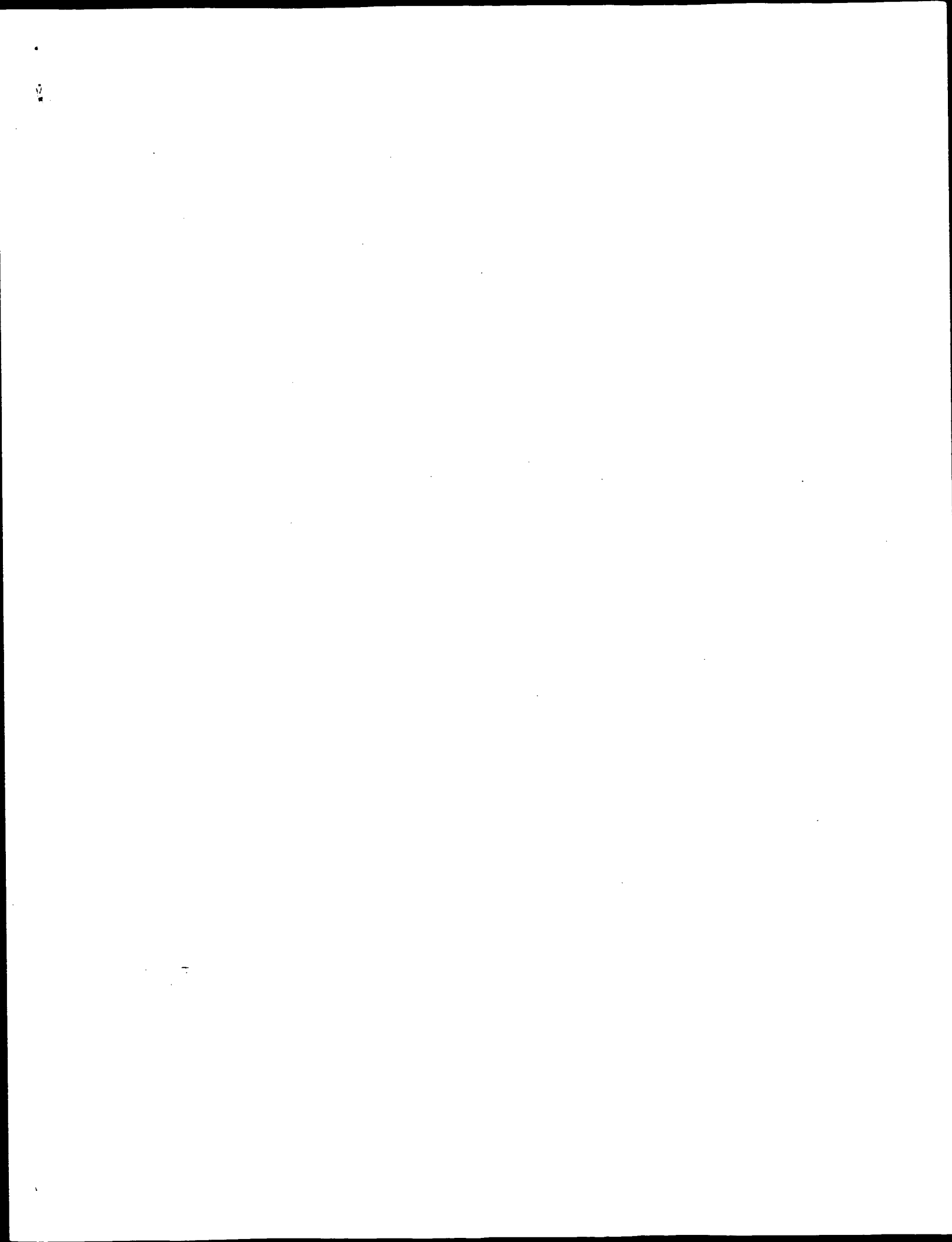
```

APPLICATION NUMBER: 08/799,910
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7853-067-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)7909090
 TELEFAX: (212)8699741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3103 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 288...1565
 OTHER INFORMATION:
 US-08-944-495-3

Query Match 76.0%; Score 15.2; DB 3; Length 3103;
 Best Local Similarity 85.0%; Pred. No. 1.3e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gatgacctcctgaccttg 20
 |||||||||||||
 Db 2031 GGTCCTCTCTGCCCCCTTG 2050

Search completed: February 24, 2002, 03:01:27
 Job time: 21438 sec



COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7333 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-2

Query Match 100.0%; Score 18; DB 4; Length 7333;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 7279 CCACAGTCGTACACCAG 7262

RESULT 3
US-08-766-528-1/c
Sequence 1, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8060 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-1

Query Match 100.0%; Score 18; DB 4; Length 8060;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 2546 CCACAGTCGTACACCAG 2529

RESULT 4
US-08-766-528-3/c
Sequence 3, Application US/08766528
Patent No. 6190861

GENERAL INFORMATION:
APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8132 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-766-528-3

Query Match 100.0%; Score 18; DB 4; Length 8132;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccacagtcgtacaccag 18
|||||
Db 8080 CCACAGTCGTACACCAG 8063

RESULT 5
US-08-766-528-61
Sequence 61, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:

APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,528
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-038CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-766-528-61

Query Match 94.4%; Score 17; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccag 18
|||||
Db 1 CACAGTCGTACACCAG 17

RESULT 6
US-09-111-085-15/c
; Sequence 15, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 28
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
US-09-111-085-15

Query Match 94.4%; Score 17; DB 3; Length 28;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccag 18
|||||
Db 28 CACAGTCGTACACCAG 12

RESULT 7
US-09-150-133-8
; Sequence 8, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820,504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-150-133-8

Query Match 76.7%; Score 13.8; DB 3; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1,1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccag 18
|||||
Db 939 cacagtcgtacaccag 955

RESULT 8
US-09-150-141-8
; Sequence 8, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820,495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-150-141-8

Query Match 76.7%; Score 13.8; DB 3; Length 1760;
Best Local Similarity 88.2%; Pred. No. 1,1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cacagtcgtacaccag 18
|||||
Db 939 cacagtcgtacaccag 955

RESULT 9
US-09-374-493-8
; Sequence 8, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:

APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-374-493-8

Query Match 76.7%: Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%: Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacaccag 18
||||| | |||||
Db 939 cacagtcgtacaccag 955

RESULT 10
US-09-374-824-8
; Sequence 8, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-374-824-8

Query Match 76.7%: Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%: Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacaccag 18
||||| | |||||
Db 939 cacagtcgtacaccag 955

RESULT 11
US-09-374-492-8
; Sequence 8, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13

EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 8
; LENGTH: 1760
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-374-492-8

Query Match 76.7%: Score 13.8; DB 4; Length 1760;
Best Local Similarity 88.2%: Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cacagtcgtacaccag 18
||||| | |||||
Db 939 cacagtcgtacaccag 955

RESULT 12
US-08-891-298-2
; Sequence 2, Application US/08891298
; Patent No. 6300488
; GENERAL INFORMATION:
; APPLICANT: Gage, Frederick H.
; APPLICANT: Suh, Steven T.
; TITLE OF INVENTION: Modified Lepidopteran Receptors
; TITLE OF INVENTION: and Hybrid Multi-Functional Proteins for Use in Transcript
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,298
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 185...2002
; OTHER INFORMATION:
US-08-891-298-2

Query Match 76.7%; Score 13.8; DB 4; Length 2711;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
DB 1092 CCACAGTCGACACACCAC 1108

RESULT 13

US-08-042-747A-4/C
; Sequence 4, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinticariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 269..2941
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..249
; US-08-042-747A-4

Query Match 76.7%; Score 13.8; DB 1; Length 3177;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccacagtcgtacaccac 18
|||||
DB 166 CGCGGTCTGACACCACG 150

RESULT 14

US-08-755-587-27

; Sequence 27, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ. ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5892 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 483..5412
; FEATURE:
; NAME/KEY: exon
; LOCATION: 481..5412
; US-08-755-587-27

Query Match 76.7%; Score 13.8; DB 3; Length 5892;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
DB 235 CCACAGCATACACACCAC 251

RESULT 15

US-08-480-784-20/c
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff

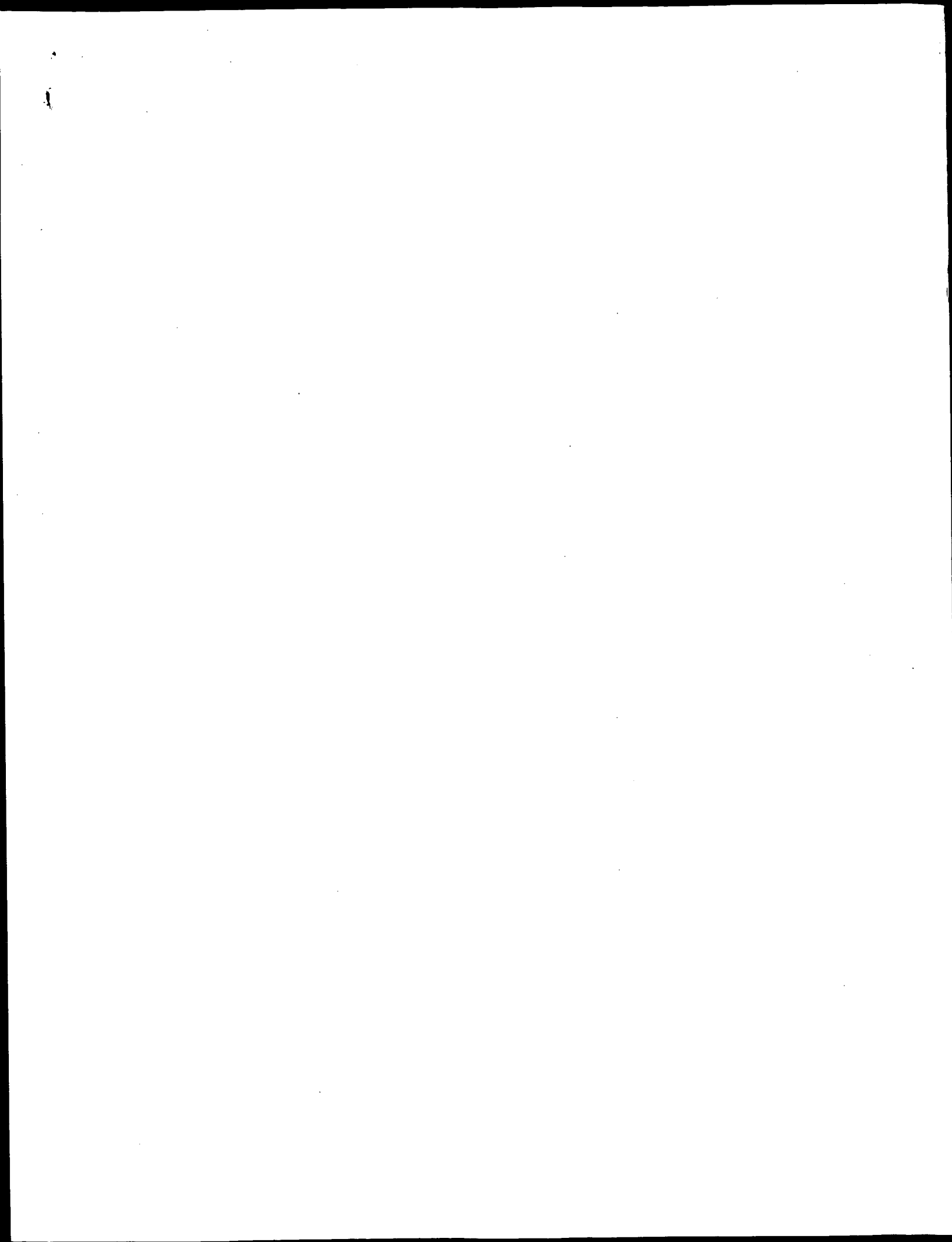
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavlignan, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,784
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ. ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-480-784-20

Query Match 76.7%; Score 13.8; DB 1; Length 6769;
Best Local Similarity 88.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
| |||| |||||
Db 965 CTACAGCGGTACACCAC 949

Mon Feb 25 07:44:10 2002

us-09-171-553b-14.feb22std.rni



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 21:47:34 ; Search time 6170.56 Seconds

(without alignments)
5781.645 Million cell updates/sec

Title: US-09-171-553B-1

Perfect score: 3320

Sequence: 1 gaattcgccgcccgtcgac.....aaaaaaaaaaaaaaaaa 3320

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vtl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	74	2.2	391	11	BF709087 MI-P-AVO-
C 2	74	2.2	545	10	BE013835 BE013835
C 3	57	1.7	399	11	BI183551 BI183551
C 4	52	1.6	403	10	AM447645 AM447645
C 5	52	1.6	451	13	AZ891915 AZ891915
C 6	50	1.5	549	13	AZ380763 AZ380763
C 7	42	1.3	576	13	AZ730616 AZ730616
C 8	40	1.2	704	13	BH059214 BH059214
C 9	39	1.2	100	11	BF442561 BF442561
C 10	37	1.1	482	13	AO920035 AO920035
C 11	36	1.1	545	10	AA529580 AA529580
C 12	36	1.1	575	13	AZ479717 AZ479717

C 13	36	1.1	583	13	A2415280 A2415280
C 14	36	1.1	599	13	A2944719 A2944719
C 15	36	1.1	610	13	A2859488 A2859488
C 16	36	1.1	640	13	A2997740 A2997740
C 17	36	1.1	646	13	A2638811 A2638811
C 18	36	1.1	655	13	A2740414 A2740414
C 19	36	1.1	660	13	A2353293 A2353293
C 20	36	1.1	673	13	A2329067 A2329067
C 21	34	1.0	345	13	A2780588 A2780588
C 22	34	1.0	406	13	A2506946 A2506946
C 23	34	1.0	550	13	A2390750 A2390750
C 24	34	1.0	574	13	A2390749 A2390749
C 25	34	1.0	630	13	A2838774 A2838774
C 26	33	1.0	396	10	AA736984 AA736984
C 27	33	1.0	402	11	B671688 B671688
C 28	33	1.0	455	13	A2818850 A2818850
C 29	32	1.0	206	13	A2792975 A2792975
C 30	32	1.0	229	13	A2914781 A2914781
C 31	32	1.0	298	10	AM102270 AM102270
C 32	32	1.0	353	10	AA918868 AA918868
C 33	32	1.0	358	11	BF604441 BF604441
C 34	32	1.0	361	10	AM074793 AM074793
C 35	32	1.0	465	13	A2431949 A2431949
C 36	32	1.0	502	13	A2447630 A2447630
C 37	32	1.0	512	13	A2459290 A2459290
C 38	32	1.0	518	10	AL499730 AL499730
C 39	32	1.0	523	13	A2995885 A2995885
C 40	32	1.0	556	11	B606618 B606618
C 41	32	1.0	561	13	A2793135 A2793135
C 42	32	1.0	561	13	A2951769 A2951769
C 43	32	1.0	567	13	A2624307 A2624307
C 44	32	1.0	568	13	A2824214 A2824214
C 45	32	1.0	573	13	A2797677 A2797677

ALIGNMENTS

RESULT 1
BF709087 391 bp mRNA EST 02-JAN-2001
LOCUS MI-P-AVO-nae-c-07-0-01.s1 MI-P-AVO Sus scrofa cDNA clone
DEFINITION MI-P-AVO-nae-c-07-0-01 3', mRNA sequence.
ACCESSION BF709087
VERSION BF709087.1 GI:12008564
KEYWORDS EST.
SOURCE Sus scrofa
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 391)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: ctugle@iastate.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized placenta library cDNA library Preparation: M.B. Soares Lab, University of Iowa
University of Iowa Clome Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward

FEATURES
Source
POLYA=yes.
Location/Qualifiers
1. .391
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="MI-P-AV0-nae-c-07-0-UU"
/lab_host="MI-P-AV0"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: EcoRI; The MI-P-AV0 library is derived from placenta. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigst.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=MI-P-AV0
TAG_TISSUE=placenta
TAG_SEQ=ATTGCG"
BASE COUNT 94 a 87 c 71 g 139 t
ORIGIN

Query Match 2.2%; Score 74; DB 11; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1984 gttttgtagacaccccttcagagatggtagagagctatcctactaagaagagactca 2043
|||||
Db 250 gttttgtagacaccccttcagagatggtagagagctatcctactaagaagagactca 191
|||||

QY 2044 accgtgtgtgctaa 2057
|||||
Db 190 accgtgtgtgctaa 177
|||||

RESULT 2
BE013835/c
LOCUS BE013835 545 bp mRNA EST 09-JUL-2000
DEFINITION 125326 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE013835
VERSION BE013835.1 GI:8274822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mainscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 52 row: B column: 23
Seq primer: ATTAGTGTGACACTATAG.
Location/Qualifiers
1. .345
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"

FEATURES
Source
/tissue_type="pooled"
/lab_host="DHI0B"
/note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 128 a 120 c 129 g 168 t
ORIGIN

Query Match 2.2%; Score 74; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1984 gttttgtagacaccccttcagagatggtagagagctatcctactaagaagagactca 2043
|||||
Db 216 gttttgtagacaccccttcagagatggtagagagctatcctactaagaagagactca 157
|||||

QY 2044 accgtgtgtgctaa 2057
|||||
Db 156 accgtgtgtgctaa 143
|||||

RESULT 3
B1183551
LOCUS B1183551 399 bp mRNA EST 10-JUL-2001
DEFINITION UNL-P-FN-by-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION B1183551
VERSION B1183551.1 GI:14657960
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Caetano,A.R., Johnson,R.K. and Pomp,D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
Oligo-dT track not found. Not I site shown in beginning of sequence
is likely internal to the message.
Seq primer: M13 -29
POLYA=No.

FEATURES
Source
Location/Qualifiers
1. .399
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-by-g-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DHI0B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_SEQ=None found"

BASE COUNT 120 a 83 c 92 g 96 t 8 others

ORIGIN

Query Match 1.7%: Score 57: DB 11: Length 399:
Best Local Similarity 100.0%: Pred. No. 1.4e-06;
Matches 57: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1920 cgtcactgggaagtgaactcactgaggaagccgagcgaataacgaacaata 1976
116 CGGCACCTGGGAGTGGACTTCACTGAGGTAAGCCGCGTAATACGGAACAATA 172

RESULT 4
AM447645 403 bp mRNA EST 25-APR-2001

LOCUS 89634 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
DEFINITION AM447645
ACCESSION AM447645.1 GI:6989432
VERSION EST.
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 403)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McCorm,C.G.,
Perta,G., Holt,L., Karanycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013

COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.960904.e. Vector identified by cross_match with the -minscore 20
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 70 row: L column: 19
Seq primer: ATTGAGTGACACTATAG.
Location/Qualifiers

FEATURES

source 1..403
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 61 a 134 c 124 g 84 t
ORIGIN

Query Match 1.6%: Score 52: DB 10: Length 403:
Best Local Similarity 100.0%: Pred. No. 3.5e-05;
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2461 gagtgggtgagcagcagcgtggaagcagctccggaaggaagcactactcagga 2512
116 ggtgggtgagcagcagcagcgtggaagcagctccggaaggaagcactactcagga 2512

RESULT 5
A2891915/c

LOCUS A2891915 451 bp DNA GSS 05-MAR-2001
DEFINITION RPCI-24-172A20.TV RPCI-24 Mus musculus genomic clone RPCI-24-172A20
, DNA sequence.
ACCESSION A2891915
VERSION A2891915.1 GI:13210860
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 451)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akınret,B., Levins,M.,
Tsegaye,G., Geier,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-172A20.TJ
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdjong@mai.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 172 row: A column: 20
Seq primer: 77
Class: BAC ends.

FEATURES

source 1..451
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-24-172A20"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 144 a 91 c 106 g 110 t
ORIGIN

Query Match 1.6%: Score 52: DB 13: Length 451:
Best Local Similarity 100.0%: Pred. No. 3.2e-05;
Matches 52: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 2420 tgcctgtccagcagcctgtctctcctgaagcagcagcagcagcagcagcagc 2471
116 tgcctgtccagcagcctgtctctcctgaagcagcagcagcagcagcagcagc 2471

Db 304 TGCTGCTTCCACAGCCTTGTTCTCTAGGCTCAAGCGCTCGAGGCGGTGAG 253
RESULT 6
A2380763 549 bp DNA GSS 02-OCT-2000
LOCUS 1M0136B22R Mouse 10kb plasmid U061M Library Mus musculus genomic
clone U061M0136B22 R, DNA sequence.
ACCESSION A2380763
VERSION A2380763.1 GI:10494463
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 549)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hambl,C.,

TITLE
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.
JOURNAL
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0136 row: B column: 22
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 549.

FEATURES

SOURCE
1. 549

Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC1M0136B22"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114[db|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

159 a 119 c 119 g 152 t

ORIGIN

Query Match
Best Local Similarity 1.5%; Score 50; DB 13; Length 549;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2131 gtctccagagtaagtcagagctgccaagatattggagttattgga 2180
|||||
Db 228 GTTCCCAAGTAAGTCAAGGACGTGGCCAAAGATATTGGGATGATGGAA 277

RESULT 7
LOCUS AZ730616
DEFINITION RPCR-24-149D1.TJ RPCR-24 Mus musculus genomic clone RPCR-24-149D1,
ACCESSION AZ730616
VERSION AZ730616
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 576)

AUTHORS
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akınret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P., and Fraser, C.M.
JOURNAL
Mouse BAC End Sequences from Library RPCR-24
Unpublished (1999)
COMMENT
Other_GSSs: RPCR-24-149D1.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCR-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choori.org/bacpac/orderframe.htm). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 149 row: D column: 1
Seq primer: SP6
Class: BAC ends.

FEATURES

source
1. 576

Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCR-24-149D1"
/clone_lib="RPCR-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI. RPCR-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT
161 a 139 c 121 g 155 t

ORIGIN

Query Match
Best Local Similarity 1.3%; Score 42; DB 13; Length 576;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2406 acatagctgctgctgctgcttccagccttctctctag 2447
|||||
Db 317 ACATAGCTGCTGCTGCTCTTCCACGCTTGTCCTAG 358

RESULT

8

LOCUS BH059214
DEFINITION RPCR-24-326H3.TJ RPCR-24 Mus musculus genomic clone RPCR-24-326H3,
ACCESSION BH059214
VERSION BH059214
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 704)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akınret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P., and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCR-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCR-24-326H3.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 326 row: H column: 3
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..704

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-326H3"
/clone_11b="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: PTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 224 a 132 c 174 g 174 t
ORIGIN

Query Match 1.2% Score 40; DB 13; Length 704;
Best Local Similarity 100.0%; Pred. No. 0.049; Mismatches 0; Indels 0; Gaps 0;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2137 caggttaagtcaggagctggccaagatattggagattgatt 2176
|||||
Db 562 CAGGTAAAGTCAGGACTGGCCAGATATTGGGCAATTGATT 601

RESULT 9
LOCUS BF442561 100 bp mRNA EST 01-DEC-2000
DEFINITION 259390 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF442561
VERSION BF442561.1 GI:11502653
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 100)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.

REFERENCE

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

JOURNAL

COMMENT

Email: smtlhem@mail.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 90 row: 1 column: 16
Seq primer: ATTAGGTGACACTATG.

FEATURES

Location/Qualifiers
1..100

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_11b="MARC 2P1G"

/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

Query Match 1.2% Score 39; DB 11; Length 100;
Best Local Similarity 100.0%; Pred. No. 0.46; Mismatches 0; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1348 ggggcatcattaaacaaagggtgtctacccaagca 1386
|||||
Db 55 GGGGCATCTATTAACAAAGCGGTGCTTACCTCAGCA 93

RESULT 10
LOCUS AO920035 482 bp DNA GSS 21-DEC-1999
DEFINITION RPCI-23-273D7.TV RPCI-23 Mus musculus genomic clone RPCI-23-273D7,
DNA sequence.
ACCESSION AO920035
VERSION AO920035.1 GI:6609038
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 482)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akiret,
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.

REFERENCE

Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

JOURNAL

COMMENT

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.html>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 273 row: D column: 7
Seq primer: 17
Class: BAC ends.

FEATURES

Location/Qualifiers
1..482

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-273D7"
/clone_11b="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 128 a 91 c 106 g 154 t 3 others
ORIGIN

Query Match 1.1% Score 37; DB 13; Length 482;

Best Local Similarity 100.0%; Pred. NO. 0.46;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2420 tgcgtcttcacagccttgcctctagctcaagc 2456
|||||
Db 29 TCGTCTTCCAGCCTTGTCTCTAGCCTCAAGC 65

RESULT 11

AA529580

LOCUS 545 bp mRNA EST 22-JUL-1997
DEFINITION v141604.f1 Beddington mouse embryonic region Mus musculus cDNA
clone IMAGE:906543 5' similar to SW:POL_GALV P21414 POL POLYPEPTIDE
; mRNA sequence.

ACCESSION AA529580
VERSION AA529580.1 GI:2272286

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 545)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Reising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Mesteron, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:527007
Seq primer: -28m3 rev1 ET from Amersham

High quality sequence stop: 326.
Location/Qualifiers

FEATURES

source

1. 545
/organism="Mus musculus"
/strain="C57BL/6 x DBA"
/db_xref="taxon:10090"
/clone="IMAGE:906543"
/clone_1ib="Beddington mouse embryonic region"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="7.5dpc"
/lab_host="DH12S"
/note="organ: whole embryo; Vector: pCMV-SPORT; Site: 1;
Site: 2; Not: Cloned unidirectionally. Primer:
Oligo dt. Gastrulating embryos were collected at 7.5dpc
from C57BL/6 x DBA matings, excluding embryos that had
developed head folds and all extremembryonic tissues.
Average insert size: 1.3 kb (range 0.5 - 3.0 kb).
Referenced in Development 121, 2479-2489 (1995)."

BASE COUNT 143 a 130 c 127 g 145 t

ORIGIN

Query Match 1.1%; Score 36; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. NO. 0.78;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2420 tgcgtcttcacagccttgcctctagctcaagc 2455
|||||
Db 176 TCGTCTTCCAGCCTTGTCTCTAGCCTCAAG 111

RESULT 12

A2479717/c

LOCUS 575 bp DNA

DEFINITION A2479717 mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC1M0300M05 R. DNA sequence.

ACCESSION A2479717
VERSION A2479717.1 GI:10639886

KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 575)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0300 Row: M Column: 05
Seq primer: CACACAGGAAACAGCTAGAAC
Class: plasmid ends
High quality sequence stop: 575.
Location/Qualifiers

FEATURES

source

1. 575
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC1M0300M05"
/clone_1ib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD24hv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD2 (g1147321149b/AT129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 136 a 140 c 137 g 162 t

ORIGIN

Query Match 1.1%; Score 36; DB 13; Length 575;
Best Local Similarity 100.0%; Pred. NO. 0.75;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2203 caaagctcagacagctagagagatgaagaacc 2218
|||||
Db 160 CAAAGCTCAGGACAGCTAGAGAGATGAAGAACC 125

RESULT 13
 AZ415280/c
 LOCUS 583 bp DNA
 DEFINITION 1M0190M06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0190M06 F, DNA sequence.
 ACCESSION AZ415280
 VERSION AZ415280.1 GI:10539293
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 583)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0190 row: M column: 06
 Seq primer: CGTTGTAACGACGCGCCGT
 Class: plasmid ends
 High quality sequence stop: 583.
 Location/Qualifiers
 1..583
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0190M06"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 146 a 149 c 130 g 158 t
 ORIGIN
 Query Match 1.1%; Score 36; DB 13; Length 583;
 Best Local Similarity 100.0%; Pred. No. 0.74;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2203 caaagctcagacagtagatgaatgaatgaac 2238
 ||||||||||||||||||||||||||||||||||||
 Db 193 CAAAGCTCAGACAGTAGAGATGAATGAAAC 158

RESULT 14
 AZ944719
 LOCUS 599 bp DNA
 DEFINITION 2M0205L23R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 clone UUGC2M0205L23 R, DNA sequence.
 ACCESSION AZ944719
 VERSION AZ944719.1 GI:13809008
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 599)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0205 row: L column: 23
 Seq primer: CACACAGCAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 599.
 Location/Qualifiers
 1..599
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0205L23"
 /clone_lib="Mouse 10kb plasmid UUGC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g11473211419b1AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 160 a 136 c 151 g 152 t
 ORIGIN
 Query Match 1.1%; Score 36; DB 13; Length 599;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2203 caaagctcagacagtagatgaatgaatgaac 2238
 ||||||||||||||||||||||||||||||||||||
 Db 334 CAAAGCTCAGACAGTAGAGATGAATGAAAC 369

RESULT 15

AZ859488

LOCUS

AZ859488 610 bp DNA

GSS 21-FEB-2001

2M0165D07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION

AZ859488

VERSION

AZ859488.1 GI:13053858

KEYWORDS

GSS.

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 610)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0165 row: D column: 07
 Seq primer: CGTTGTAACGACGCGCCAGT
 Class: Plasmid ends
 High quality sequence stop: 610.
 Location/Qualifiers

FEATURES

source

1..610
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0165D07"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42ny; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g11473211419b/AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT

194 a 145 c 130 g 141 t

ORIGIN

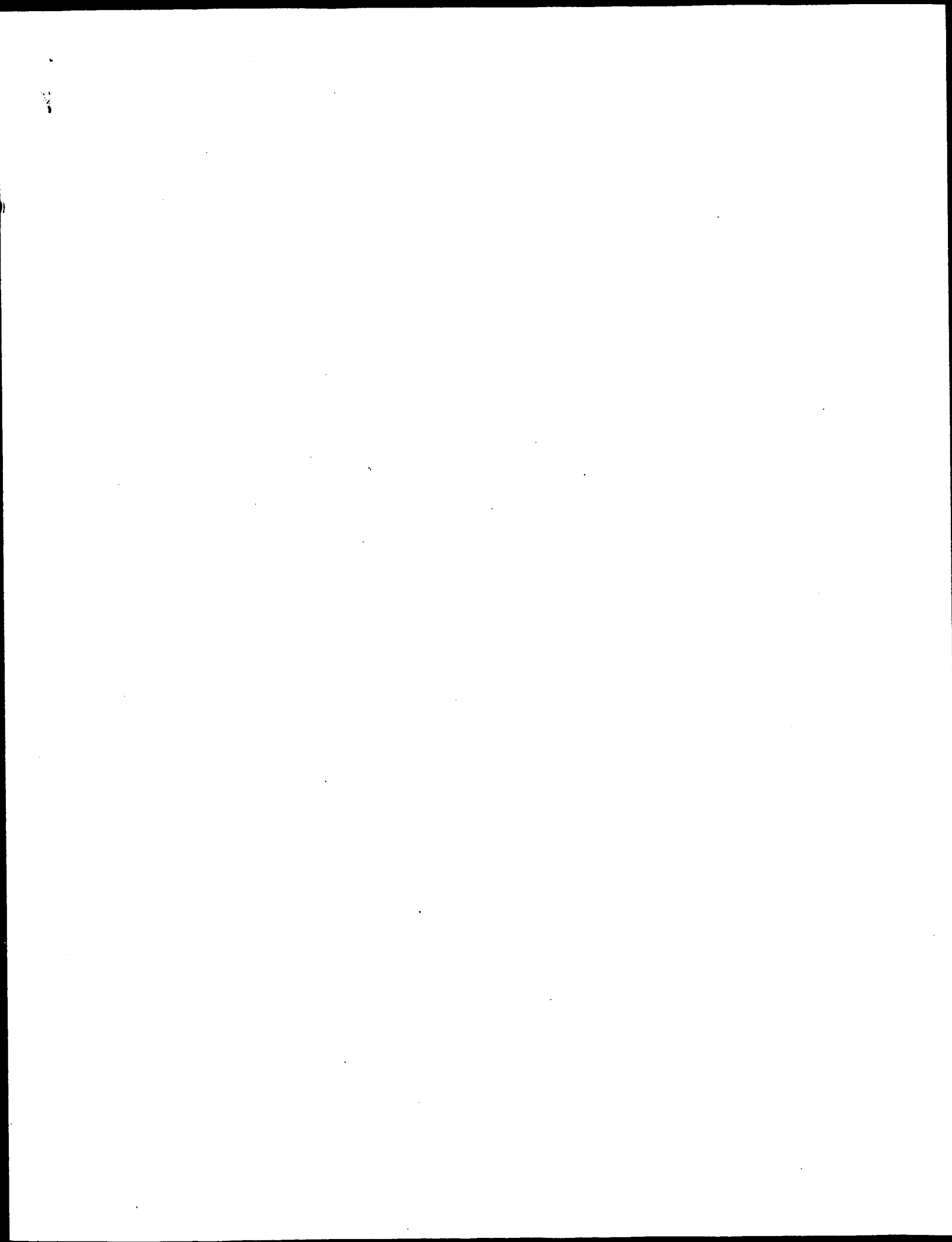
Query Match 1.1%; Score 36; DB 13; Length 610;
 Best Local Similarity 100.0%; Pred. No. 0.71;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2420 tgcgtctccagccttctctctagctcaagg 2455

|||||

DB 410 TGCCTCTTCCAGCCTTTGTTCTCTAGGCTCAAG 445

Search completed: February 24, 2002, 04:28:16
 Job Time: 24042 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:16 ; Search time 6170.56 Seconds
(without alignments)
14295.639 Million cell updates/sec

Title: US-09-171-553B-3

Perfect score: 8209

Sequence: 1 gtggtgctacgactgtggtcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	190	2.3	564	11	BI182930 UNL-P-FN-
2	140	1.7	466	11	BF441468 257862 MA
3	140	1.7	467	11	BF441465 257856 MA
C 4	130	1.6	532	11	BF709893 MI-P-A11-
5	126	1.5	408	11	BF711018 MI-P-A11-
C 6	102	1.2	564	11	BI181018 MI-P-A11-
7	102	1.2	564	11	BI181099 UNL-P-FN-
8	96	1.2	376	11	BI186129 UNL-P-FN-
9	92	1.1	558	11	BF441469 257863 MA
C 10	92	1.1	602	11	BI360879 388957 MA
11	80	1.0	452	11	BF712162 MI-P-F6-a
12	78	1.0	549	10	BF441466 257857 MA
					AW657531 110922 MA

13	77	0.9	857	11	BI184013 UNL-P-FN-
C 14	74	0.9	391	11	BF709087 MI-P-AV0-
C 15	74	0.9	545	10	BE013835 125326 MA
C 16	72	0.9	229	11	BF443400 260872 MA
C 17	69	0.8	468	11	BI186066 UNL-P-FN-
C 18	69	0.8	499	11	BF704354 MI-P-O3-a
C 19	69	0.8	521	11	BI399852 MI-P-AV1-
C 20	69	0.8	730	11	BI185535 UNL-P-FN-
C 21	68	0.8	468	11	BI182727 UNL-P-FN-
C 22	64	0.8	447	11	BI401063 MI-P-CP0-
C 23	64	0.8	485	11	BF703671 MI-P-E4-a
C 24	64	0.8	554	11	BF703805 MI-P-O2-a
C 25	63	0.8	575	11	BI183356 UNL-P-FN-
C 26	59	0.7	465	11	BI399234 MI-P-AV1-
C 27	57	0.7	399	11	BI183551 UNL-P-FN-
C 28	55	0.7	481	10	AM231947 2C22N1 SU
C 29	52	0.6	403	10	AM447645 89634 MAR
C 30	52	0.6	451	13	A2891915 RPCI-24-1
C 31	51	0.6	545	11	BF701947 MI-P-E3-a
C 32	50	0.6	494	11	BI182742 UNL-P-FN-
C 33	50	0.6	549	13	A2380763 1M0136B22
C 34	49	0.6	300	11	BI184279 UNL-P-FN-
C 35	48	0.6	345	10	AM358862 43799 MAR
C 36	48	0.6	595	11	BI185465 UNL-P-FN-
C 37	47	0.6	294	11	BF713588 MI-P-O2-a
C 38	47	0.6	579	11	BI183723 UNL-P-FN-
C 39	45	0.5	286	13	A2771375 1M0573C02
C 40	45	0.5	575	13	BH025657 RPCI-24-3
C 41	45	0.5	599	13	A2041781 RPCI-23-2
C 42	44	0.5	173	10	AM435835 75099 MAR
C 43	44	0.5	341	10	AM416859 52601 MAR
C 44	42	0.5	576	13	A2730616 RPCI-24-1
C 45	41	0.5	256	10	AM308385 3994 MARC

ALIGNMENTS

RESULT 1
BI182930/c
LOCUS UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
DEFINITION UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
BI182930
ACCESSION BI182930.1 GI:14657339
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Caetano, A.R., Johnson, R.K. and Pomp, D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized porcine ovarian follicles library
Seq primer: M13 -29
POLYA-Yes

FEATURES
source
Location/Qualifiers
1..564
/organism="Sus scrofa"

FAX: 402 / 62 4390

FEATURES

	Seq primer:	Location/Qualifiers
Plate:	86	row: C column: 8
	ATTTAGGTGACACTATAG.	

```
Location/Qualifiers
1..532
/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MT-P-A11-ngsa-b-06-0-01"
/clone_1lb="MT-P-A11"
/lab_host="BHI0B (Life Technologies)"
/notes="Vector: pR73D-Pac (Pharmacia) with a modified
```

```

FEATURES
  source
    elements were found in this CDS
    >GC_richlow_complexity
    Seq primer: M13 Forward
    POLYA-No.
    Location/Qualifiers
      1..408
        /organism="Sus scrofa"
        /strain="Grosdbreed"
        /db_xref="taxon:9623"

```

```

/clone="MI-P-AV1-ndp-h-02-0-UT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-AV1
library is normalized library derived from the MI-P-AV0
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigst.genome.iastate.edu/. The procedure used to
create this library has been previously described (Bonaldo
, Lennon and Soares, Genome Research 6: 791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      94 a 139 c 122 g 52 t 1 others
ORIGIN
Query Match      1.5%; Score 126; DB 11; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.6e-22;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 419 ctgcgtcagagaccagctctgttgaaagaaagctcccccgcgcgcgcgc 478
|||||
Db 372 CTGTCGTGACAGACCGAGTCTGTGTTGAAGCGAAAGCTTCCCTCCGCGCCGTC 313
QY 479 cgaactcttgctgtgttgaaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 538
|||||
Db 312 CGACTCTTTGCTGCTGTGGAAGACGCGACGCGTGCCTGTGATCTGTGTGT 253
QY 539 tcttgt 544
Db 252 TTCTGT 247

RESULT 6
LOCUS      B1181099      564 bp      mRNA      EST      10-JUL-2001
DEFINITION UNL-P-FN-ab-d-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION  B1181099
VERSION     B1181099
KEYWORDS   EST.
SOURCE     B1181099.1 GI:14655508
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 564)
AUTHORS   Caetano, A.R., Johnson, R.K. and Pomp, D.
TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
JOURNAL   Unpublished (2001)
COMMENT   Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found. Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA-No.

FEATURES
Source
Location/Qualifiers
1..564
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-ab-d-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified

```

```

polylinker. Site 1: Not I; Site 2: Eco RI. The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
Oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.
TAG_SEQ=None found"
BASE COUNT      112 a 196 c 143 g 112 t 1 others
ORIGIN
Query Match      1.2%; Score 102; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1204 gccaatgagcccccagatattgccccttctctctgcagatctataatgaaana 1263
|||||
Db 372 GCCAATTGACAGCCCGCCGATATGCGCCCTTTCTCTGCAGATCTATATGGAANA 431
QY 1264 ctaccatccccccttctcgagagatcccaacgcctcacgg 1305
|||||
Db 432 CTAAACCATCCCGCTTCTCGAGAGATCCCAACGCTCACGG 473

RESULT 7
LOCUS      B1186129      564 bp      mRNA      EST      10-JUL-2001
DEFINITION UNL-P-FN-ab-d-02-0-UNL.s2 UNL-P-FN Sus scrofa cDNA clone
ACCESSION  B1186129
VERSION     B1186129
KEYWORDS   EST.
SOURCE     B1186129.1 GI:14660538
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 564)
AUTHORS   Caetano, A.R., Johnson, R.K. and Pomp, D.
TITLE     Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
JOURNAL   Unpublished (2001)
COMMENT   Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpomp@unl.edu
            Oligo-dT track not found. Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA-No.

FEATURES
Source
Location/Qualifiers
1..564
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-ab-d-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI. The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6

```

nucleotides present between the Not I site and the oligo-dr track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 112 a 196 c 144 g 112 t
TAG_SEQ=None found"

Query Match 1.2%; Score 102; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 gccaatgagagcccccagatgagccttctctcgcagatccttaattgaana 1263
|||||
DB 372 GCCAATGAGAGCCCTTCACATATTGAGCCCTTCTCTGACATCTAATATGGAAAA 431

QY 1264 ctaccatccccctctctcgcagatcccccaagcctcaag 1305
|||||
DB 432 CTACCATCCCTTCTCTGAGAGATCCCAAGCCTCAGG 473

RESULT 8
LOCUS BF441469 376 bp mRNA EST 01-DEC-2000
DEFINITION 257663 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF441469
VERSION BF441469.1 GI:11501633
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 376)
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 86 row: D column: 9
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 376

FEATURES
source

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 91 a 94 c 108 g 83 t
ORIGIN

Query Match 1.2%; Score 96; DB 11; Length 376;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1359 cagctgtcgacacactctcaacacgagagagagaattctgttagaggtcaga 1418

|||||
DB 61 CAGCTGTGACAGACACTCTTCACACACGAGAGAGAGAAATCTGTAGAGCTAGA 120

QY 1419 aaaatgtctctgagccgcagagcgagaccacgcag 1454
|||||

DB 121 AAAAATGTTCCGTGGGGCGGAGCGGCGGACCGACGCGAG 156

RESULT 9
LOCUS B1360879 558 bp mRNA EST 01-AUG-2001
DEFINITION 388957 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION B1360879
VERSION B1360879.1 GI:15056907
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 558)
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross-match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 142 row: B column: 4
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 558

FEATURES
source

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 129 a 123 c 119 g 187 t
ORIGIN

Query Match 1.1%; Score 92; DB 11; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.6e-14;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8098 gactcggagcgagagctctcaccgcgagtggtgtaagactgtggagccagcgct 8157
|||||

DB 132 GCACCTGGGGCGGCGAGTCTCTACCCCTGCGGTGAGACTGTGGCCGCGCGCT 191

QY 8158 tgaataaaaatccctctctgtctgttgatcaaa 8189
|||||

DB 192 TGGAAATAAAAATCTCTTCTGCTTTTGCAATCAA 223

RESULT 10
LOCUS BF712162 602 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-E6-acx-e-11-1-UM.s1 MI-P-E6 Sus scrofa cDNA clone
ACCESSION MI-P-E6-acx-e-11-1-UM 3', mRNA sequence.
BF712162

VERSION BF712162.1 GI:12011595
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 602)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugale CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: cktugale@iastate.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized fetus at gestational day 20 library cDNA library
 Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
 Research Center, Department of Animal Science, University of
 Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLY-A=yes.

FEATURES
 source
 Location/Qualifiers
 1..602
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-E6-acc-e-11-1-UM"
 /clone_lib="MI-P-E6"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTRF3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: EcoRI; The MI-P-E6
 library is derived from fetus at gestational day 20. For
 a detailed description of the library from which this
 clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_LIB=MI-P-E6
 TAG_TISSUE=fetus at gestational day 20
 TAG_SEQ=AGGAA"

BASE COUNT 141 a 146 c 152 g 162 t 1 others
 ORIGIN

Query Match 1.1%; Score 92; DB 11; Length 602;
 Best Local Similarity 99.3%; Pred. No. 4.3e-14;
 Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8060 ggccttgctgtaaacccataaagctgtccgattccgactcgggagccagctcct 8119
 |||||||
 DB 143 GGCCTTGCTGTGAACCCATTAAGAGCTGCCGATTCGCCGACGCGGCGCGAGTCTCT 84
 |||||||

QY 8120 acccctgcgtggtgtaagctgtgggccccagcgccgttggaataaactcctctgcgcg 8179
 |||||||
 DB 83 ACCCCTGCCTGCGTACGACTGTGGGCCCGCCAGCGCGCTTGGAATTAATAATCTCTGCG 24
 |||||||

QY 8180 ttgtcatcaaaaaaaaaaaaaa 8202
 |||||||
 DB 23 TTTCATCATCAAAAAAAAAAAAA 1
 |||||||

RESULT 11
 BF441466 452 bp MRNA EST 01-DEC-2000
 LOCUS
 DEFINITION 257857 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BF441466
 VERSION BF441466.1 GI:11501630
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 452)
 AUTHORS Fahrnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
 and Keele,J.W.
 TITLE Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 JOURNAL Unpublished (2000)
 COMMENT Contact: Smith RPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402/762 4366
 Fax: 402/762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -mismatch 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTACAGAG
 Plate: 86 row: C column: 9
 Seq primer: ATTTCGTCGACATCTAG.

FEATURES
 source
 Location/Qualifiers
 1..452
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 107 a 118 c 125 g 102 t
 ORIGIN

Query Match 1.0%; Score 80; DB 11; Length 452;
 Best Local Similarity 100.0%; Pred. No. 5.8e-11;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1419 aaaatgttcctggggccgacggcgaccccgagtgcaaaatgagatgacatgga 1478
 |||||||
 DB 120 AAAAATGTTCTGGGGCGACGGCGACCCAGCAGTTGCAAAATGAGATTGACATGGGA 179
 |||||||

QY 1479 ttcccttgactgcgcccg 1498
 |||||||

DB 180 TTTCCTTGACATCGGCCCG 199
 |||||||

RESULT 12
 AM657531 549 bp MRNA EST 25-APR-2001
 LOCUS
 DEFINITION 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM657531
 VERSION AM657531.1 GI:7423429
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Pette,G., Holt,I., Karanymcheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and mismatch 12 options.

FEATURES
Source
1. 549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 145 a 144 c 121 g 139 t

ORIGIN

Query Match 1.0%; Score 78; DB 10; Length 549;
Best Local Similarity 99.2%; Pred. No. 1.6e-10;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6640 aaatcaatgtgacaaagagcatagaatcaatgacatgaggggtccgaaataagcgt 6699
159 AAATTCATATGTGACCAAGAGCATAGAAATCAATGTACATGGGGGCCCAATAAGCTTC 218

Db 6700 accctcaatgaatttcgggaagagcatgacatgaggaagctcccatccaccacaa 6759
219 ACCCTCACTGAAGTTTCCGGGAAGGACATGATAGGAAAGCTCCCATCCACACAA 278

QY 6760 cacccttgc 6768
279 CACCTTTGC 287

RESULT 13
LOCUS B1184013 857 bp mRNA EST 10-JUL-2001
DEFINITION UNL-P-FN-cf-b-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION B1184013
VERSION B1184013
KEYWORDS EST.
SOURCE Sus scrofa
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 857)
Caetano, A.R., Johnson, R.K. and Pomp, D.
Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
Unpublished (2001)
Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu

FEATURES
Source
1. 857
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection Lines"
/db_xref="taxon:9823"
/clone_lib="UNL-P-FN-cf-b-12-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73B-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 152 a 217 c 248 g 236 t 4 others

ORIGIN

Query Match 0.9%; Score 77; DB 11; Length 857;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 gcttgaataaatactctctgttgcataagaccgtctcgtgagtgattgg 87
38 GCTTGGAATMAAAATCTCTGCTGTCATCAAGACCGTTTCGTGAGATTGGG 97

Db 88 gttgcgcctctcgtgag 104
98 GTTGCCTCTTCGAG 114

RESULT 14
LOCUS BF709087/c 391 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-AV0-nae-c-07-0-UI.s1 MI-P-AV0 Sus scrofa cDNA clone
ACCESSION BF709087
VERSION BF709087
KEYWORDS EST.
SOURCE Sus scrofa
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 391)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
Contact: Tuglie CR
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: crtuglie@iastate.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the

non-normalized placenta library cDNA library preparation: M.B. Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab, University of Iowa Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1. 391
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="MI-P-Ay0-nae-c-07-0-UI"
/lab_host="MI-P-Ay0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: EcoRI; The MI-P-Ay0 library is derived from placenta. For a detailed description of the library from which this clone was derived, please visit our web site at http://piglet.genome.iastate.edu/. The procedure used to create this library has been previously described (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB="MI-P-Ay0"
TAG_SEQ="ATTGG"

BASE COUNT

94 a 87 c 71 g 139 t

ORIGIN

Query Match 0.9%; Score 74; DB 11; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4935 gttttgtacaccccttcagatggtagagcttacctactaagaagacttca 4994
|||||
DB 250 GTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCTACTAAGAAAGACTTCA 191
|||||
OY 4995 accgtgtggtctaa 5008
|||||
DB 190 ACCGTGTGCTTAA 177

RESULT 15
BE013835/c 545 bp mRNA EST 09-JUL-2000
LOCUS 125326 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BE013835
ACCESSION BE013835
VERSION BE013835.1 GI:8274822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 545)
AUTHORS Fahrnenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, U.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
unpublished (2000)
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.960904.e. Vector identified by crossmatch with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCA
BACKWARD: GTTTCACGACGACG
Plate: 52 row: B column: 23
Seq primer: ATTAGGTGACACTATAG.

FEATURES

source

Location/Qualifiers
1. 545
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPOR6; Site 1: XbaI; Site 2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT

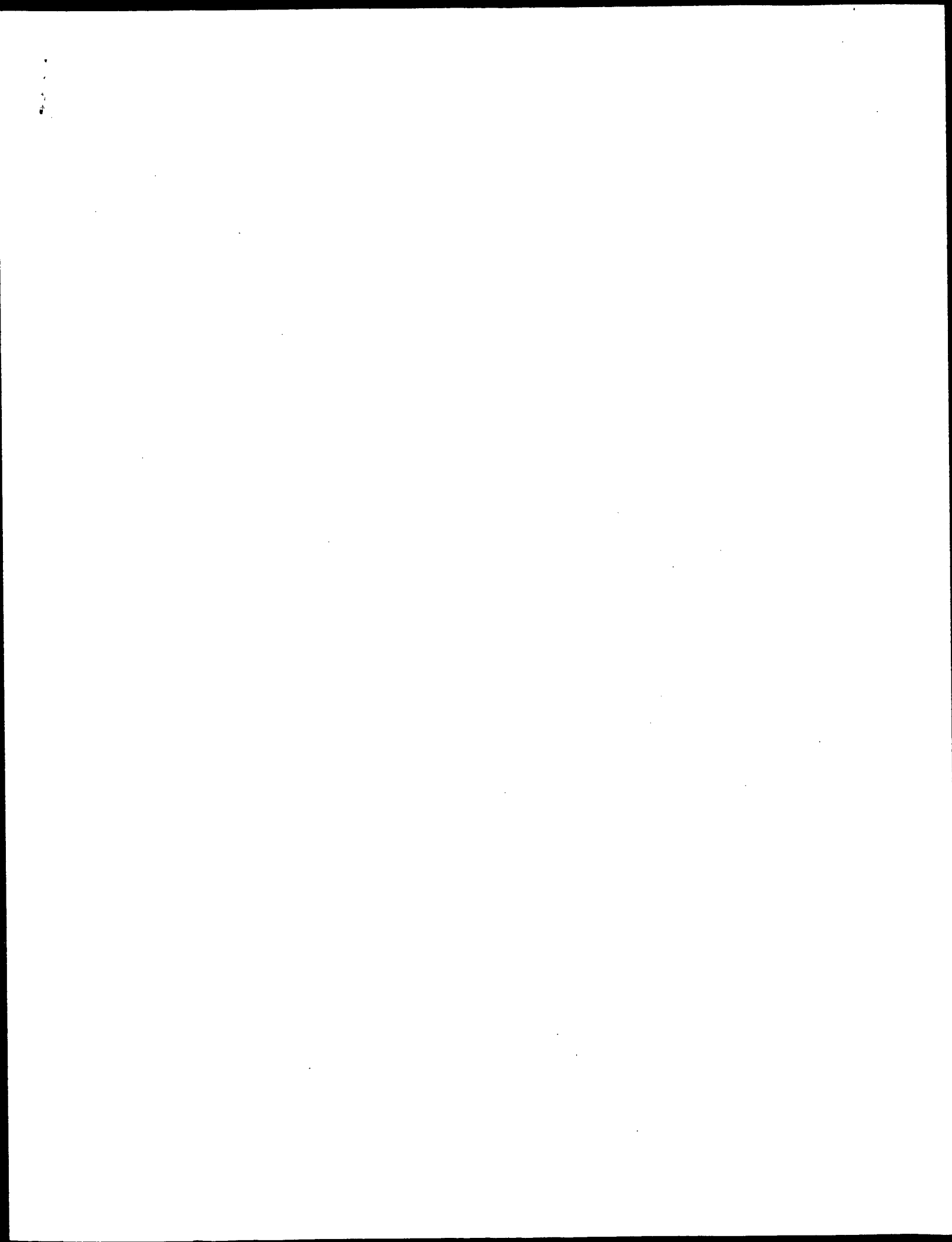
128 a 120 c 129 g 168 t

ORIGIN

Query Match 0.9%; Score 74; DB 10; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4935 gttttgtacaccccttcagatggtagagcttacctactaagaagacttca 4994
|||||
DB 216 GTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCTACTAAGAAAGACTTCA 157
|||||
OY 4995 accgtgtggtctaa 5008
|||||
DB 156 ACCGTGTGCTTAA 143

Search completed: February 24, 2002, 04:28:43
Job time: 24069 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2002, 04:28:43 ; Search time 6170.56 Seconds
(without alignments)
5147.753 Million cell updates/sec

Title: US-09-171-553b-9

Perfect score: 2956

Sequence: 1 tgcctttaggggttagaac.....aaaaaaaaaaaaaaaaaaaaa 2956

Scoring table:

OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_estc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vtl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	190	6.4	564	11	B1182930 UNL-P-FN-
C 2	92	3.1	558	11	B1360879 388957 MA
C 3	92	3.1	602	11	BF712162 MI-P-E6-a
C 4	78	2.6	549	10	AM657531 110922 MA
C 5	69	2.3	468	11	B1186066 UNL-P-FN-
C 6	69	2.3	499	11	BF704354 MI-P-O3-a
C 7	69	2.3	521	11	B1399852 MI-P-AV1-
C 8	69	2.3	730	11	B1185535 UNL-P-FN-
C 9	68	2.3	468	11	B1182727 UNL-P-FN-
C 10	64	2.2	485	11	BF703671 MI-P-E4-a
C 11	64	2.2	554	11	BF703805 MI-P-O2-a
C 12	52	1.8	451	13	A2891915 RPCI-24-1

C 13	51	1.7	545	11	BF701947 MI-P-E3-a
C 14	49	1.7	300	11	B1184279 UNL-P-FN-
C 15	48	1.6	494	11	B1182742 UNL-P-FN-
C 16	48	1.6	595	11	B1185465 UNL-P-FN-
C 17	47	1.6	294	11	BF713588 MI-P-O2-a
C 18	47	1.6	576	11	A2730616 RPCI-24-1
C 19	47	1.6	579	11	B1183723 UNL-P-FN-
C 20	40	1.4	166	10	AM346655 29171 MAR
C 21	40	1.4	631	11	BF702741 MI-P-E3-a
C 22	38	1.3	403	10	AM447645 89634 MAR
C 23	37	1.3	482	13	AQ920035 RPCI-23-2
C 24	36	1.2	545	10	AA529580 V14104.r
C 25	36	1.2	610	13	A2859488 2M0152P15
C 26	36	1.2	640	13	A2997740 2M0284R23
C 27	36	1.2	646	13	A2638811 1M0498123
C 28	36	1.2	655	13	A2740414 RPCI-24-1
C 29	36	1.2	660	13	A2353293 1M0092E07
C 30	36	1.2	673	13	A2329067 1M0053610
C 31	35	1.2	857	11	B1184013 UNL-P-FN-
C 32	34	1.2	155	10	AM346676 29231 MAR
C 33	34	1.2	261	11	B1184980 UNL-P-FN-
C 34	34	1.2	345	13	A2780588 2M0018A11
C 35	34	1.2	513	10	AM786160 119054 MA
C 36	34	1.2	550	13	A2390750 1M0152P16
C 37	34	1.2	574	13	A2390749 1M0152P15
C 38	32	1.1	340	13	BH068193 RPCI-24-3
C 39	32	1.1	348	13	A2966390 2M0236A21
C 40	31	1.0	251	11	BG382013 297673 MA
C 41	31	1.0	257	10	AA245675 mx05911.r
C 42	31	1.0	432	11	B1081943 602877172
C 43	31	1.0	834	13	BH124666 RPCI-24-2
C 44	30	1.0	114	11	BF728989 1000070E0
C 45	30	1.0	230	10	AM786010 117582 MA

ALIGNMENTS

RESULT 1
B1182930 564 bp mRNA EST 10-JUL-2001
UNL-P-FN-dr-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
B1182930
UNL-P-FN-dr-h-08-0-UNL.3', mRNA sequence.

ACCESSION B1182930.1 GI:14657339

VERSION B1182930.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: drompeunl.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library

Seg primer: M13 -29
POLA+yes.

Location/Qualifiers
1..564
/organism="Sus scrofa"

```

/strain="University of Nebraska, Lincoln Swine Selection
lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-DI-h-08-0-UNL"
/clone_11b="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dt track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.
TAG_LIB=UNL-P-FN
TAG_TISSUE=porcine ovarian follicles
TAG_SEQ=CACACT"
TAG_SEQ=CGACACT"
BASE COUNT      141 a      129 c      143 g      151 t
ORIGIN

```

```

Query Match      6.4%; Score 190; DB 11; Length 564;
Best Local Similarity 99.3%; Pred. No. 5.4e-37;
Matches 290; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2467 cctctgctgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 2526
DB 564 cctctgctgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 505
QY 2527 tctctgctcgaagtaagtaagtagtagaagtcacacttccattgttccagggcctgcta 2586
DB 504 tctctgctcgaagtaagtaagtagtagaagtcacacttccattgttccagggcctgcta 445
QY 2587 aactgactgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 2646
DB 444 aactgactgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 385
QY 2647 ctgcaatctgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 2706
DB 384 ctgcaatctgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 325
QY 2707 aatgattgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 2758
DB 324 aatgattgctcgaagtaagtagtagaagtcacacttccattgttccagggcctgcta 273

```

```

RESULT 2
LOCUS B1360879 558 bp mRNA EST 01-AUG-2001
DEFINITION B1360879 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION B1360879
VERSION B1360879.1 GI:15056907
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 558)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keeler,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
JOURNAL
COMMENT

```

```

Email: smith@emall.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACACGATGATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 142 row: B column: 4
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. 558
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_11b="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: PCMV SPORF6; Site 1: XbaI; Site 2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

```

```

BASE COUNT      129 a      123 c      119 g      187 t
ORIGIN

```

```

Query Match      3.1%; Score 92; DB 11; Length 558;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2845 gcaatcgagggccgagctctcctacccctgctgctgtagtagaagtcagacttgaggcccaagcagct 2904
DB 132 gcaatcgagggccgagctctcctacccctgctgctgtagtagaagtcagacttgaggcccaagcagct 191
QY 2905 tggaaataaatacctctctgctgctgtagtagaagtcagacttgaggcccaagcagct 2936
DB 192 tggaaataaatacctctctgctgctgtagtagaagtcagacttgaggcccaagcagct 223

```

```

RESULT 3
LOCUS BF712162/c 602 bp mRNA EST 02-JAN-2001
DEFINITION BF712162/c MI-P-E6-ax-e-11-1-UM-s1 MI-P-E6 Sus scrofa cDNA clone
ACCESSION BF712162
VERSION BF712162.1 GI:12011595
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 602)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to verify it as a clone from the
non-normalized fetus at gestational day 20 library cDNA library
preparation. RJ Woods, JA Green, RS Prather S142 Animal Science
Research Center, Department of Animal Science, University of
Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

```

FEATURES

source

Location/Qualifiers
1..602

/organism="Sus scrofa"

/strain="crossbred"

/db_xref="taxon:9823"

/clone="MI-P-E6-ack-e-11-1-UM"

/lab_host="MI-P-E6"

/note="Vector: PT73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: EcoRI. The MI-P-E6

library is derived from fetus at gestational day 20. For

a detailed description of the library from which this

clone was derived, please visit our web site at

http://pigest.genome.iastate.edu/.

TAG_LIB=MI-P-E6

TAG_TISSUE=fetus at gestational day 20

TAG_SEQ=AGGAA"

BASE COUNT 141 a 146 c 152 g 162 t 1 others

ORIGIN

Query Match

Best Local Similarity 99.3%; Pred. No. 2.1e-13;

Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2807 ggccttggtggaacccataaagctgtccgattccgacccgagccagctcct 2866

Db 143 ggcttgggtggaacccataaagctgtccgattccgacccgagccagctcct 84

QY 2867 accctcggtgtgtgaactgtgtgcccagcgctgtgaataaatacctctgtc 2926

Db 83 accctcggtgtgtgaactgtgtgcccagcgctgtgaataaatacctctgtc 24

QY 2927 ttggcatcaaaaaa 2949

Db 23 ttggcatcaaaaaa 1

RESULT 4

AM657531 549 bp mRNA EST 25-APR-2001

LOCUS 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.

AM657531

VERSION AM657531.1 GI:7423429

KEYWORDS EST.

SOURCE EST.

ORGANISM Bos taurus

COMMENT

REFERENCE

AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahnenkrug, S.C., Bennett,

G.L., Heaton, M.P., Laegreid, W.W., Kohrer, G.A., Chitko-Mckown, C.G.,

Pertea, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and

Keeler, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

CONTACT: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCACGTCACGACG

Plate: 94 row: E column: 12

FEATURES

source

Seq primer: ATTTAGTGACATATAG.

Location/Qualifiers
1..549

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pcMW SPORT6; Site 1: XbaI; Site 2: XhoI;

library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

BASE COUNT 145 a 144 c 121 g 139 t

ORIGIN

Query Match 2.6%; Score 78; DB 10; Length 549;
Best Local Similarity 99.2%; Pred. No. 5.4e-10;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1387 aattcaatgtaccacaagaagacatagaatacatgtacatggtgtccgaataagct 1446

Db 159 aattcaatgtaccacaagaagacatagaatacatgtacatggtgtccgaataagct 218

QY 1447 accctactgaagtttcgggaaggagacatgcatagagaagaagctcccatccaccaa 1506

Db 219 accctactgaagtttcgggaaggagacatgcatagagaagaagctcccatccaccaa 278

QY 1507 caactttgc 1515

Db 279 caactttgc 287

RESULT 5

BI186066 468 bp mRNA EST 10-JUL-2001

LOCUS 110922 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.

BI186066

VERSION BI186066.1 GI:14660475

KEYWORDS EST.

SOURCE EST.

ORGANISM Sus scrofa

COMMENT

REFERENCE

AUTHORS

Caetano, A.R., Johnson, R.K. and Pomp, D.

Generation and sequence characterization of a normalized cDNA

library from swine ovarian follicles

unpublished (2001)

CONTACT: Pomp, D

Department of Animal Science

University of Nebraska, Lincoln

Lincoln, NE 68583-0908, USA

Tel: 402 472 6416

Fax: 402 472 6362

Email: dpomp@unl.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail.

Seq primer: M13 -29

POLYA=Yes.

Location/Qualifiers

1..468

/organism="Sus scrofa"

/strain="University of Nebraska, Lincoln Swine Selection

lines"

/db_xref="taxon:9823"

/clone="UNL-P-FN-bo-a-05-0-UNL"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UML-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UML-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

BASE COUNT 127 a 98 c 97 g 146 t
TAG_SEQ=None found"

ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 468;
Best Local Similarity 100.0%; Pred. No. 9.3e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2884 gactgtggcccgagcgcttggaataaatacctctgtctgttgcatacaaaaaa 2943
|||||
Db 69 GACTGTGGCCCGCAGCGCTTGGAATAAATCCTCTGCTTGGCATCAAAAAA 10
QY 2944 aaaaaaaa 2952
|||||
Db 9 AAAAAAAA 1

RESULT 6
BF704354/c 499 bp mRNA EST 22-DEC-2000
LOCUS
DEFINITION MI-P-03-aba-g-01-1-UM.s1 MI-P-03 Sus scrofa cDNA clone
ACCESSION BF704354
VERSION BF704354.1 GI:11989756
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 499)

TITLE Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: RJ Woods, JA Green, RS Praether S142
Animal Science Research Center, Department of Animal Science,
University of Missouri-Columbia, 65211 Clone distribution: clones
will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source Location/Qualifiers

1..499
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-03-aba-g-01-1-UM"
/clone_lib="MI-P-03"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-03
library is derived from ovary at estrus day 12. For a

detailed description of the library from which this clone
was derived, please visit our web site at
<http://pigst.genome.iastate.edu/>.
TAG_SEQ=None found"

BASE COUNT 132 a 104 c 107 g 152 t 4 others

ORIGIN

Query Match 2.3%; Score 69; DB 11; Length 499;
Best Local Similarity 100.0%; Pred. No. 8.8e-08;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2884 gactgtggcccgagcgcttggaataaatacctctgtctgttgcatacaaaaaa 2943
|||||
Db 69 GACTGTGGCCCGCAGCGCTTGGAATAAATCCTCTGCTTGGCATCAAAAAA 10
QY 2944 aaaaaaaa 2952
|||||
Db 9 AAAAAAAA 1

RESULT 7
B1399852/c 521 bp mRNA EST 14-AUG-2001
LOCUS
DEFINITION MI-P-AV1-nrv-d-03-0-UI.s1 MI-P-AV1 Sus scrofa cDNA clone
ACCESSION B1399852
VERSION B1399852.1 GI:15178913
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 521)

TITLE Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized placenta library cDNA library Preparation: M.B. Soares
lab, University of Iowa Est sequencing: M.B. Soares lab, University
of Iowa Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source Location/Qualifiers

1..521
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-AV1-nrv-d-03-0-UI"
/clone_lib="MI-P-AV1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AV1
library is normalized library derived from the MI-P-AV1
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
<http://pigst.genome.iastate.edu/>. The procedure used to
create this library has been previously described (Bonaldo

```

, Lennon and Soares, Genome Research 6: 791-806, 1996
TAG_LIB=MI-p-A11
TAG_TISSUE=placenta
TAG_SEQ=ATTG"
BASE COUNT      142 a      106 c      103 g      170 t
ORIGIN

```

Query Match	2.3%;	Score 69;	DB 11;	Length 521;
Best Local Similarity	100.0%;	Pred. No. 8.4e-08;		
Matches 69;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 2884 gactgtgagcccccagcgcgcttggaaataaacctccttgctgctttgcacatcaaaaaaaa 2943
|||||
Dd 69 GACTGTGGCCCCACGCGCCTTGAATAAAAATCCCTTGCTTTGCATCAAAAAAAAAA 10

QY	2944	aaaaaaaaa	2952
Db	9	AAAAAAAAA	1

RESULT	8
B1185535	
LOCUS	B1185535
DEFINITION	B1185535
UNL-P-FN-cx-a-05-0-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone.	
ACCESSION	B1185535
VERSION	B1185535.1 GI:1465944
KEYWORDS	EST.
SOURCE	pig.
ORGANISM	Sus scrofa

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 730)	Caetano, A.R., Johnson, R.K. and Pomp, D.	Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles	Unpublished (2001)	Contact: Pomp, D

Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
Oligo-dr track not found, Not I site shown in beginning of sequence
is likely internal to the message.
Seq primer: M13 -29
POLYA=No.

FEATURES
SOURCE

```
Location/Qualifiers
1. .730
   /organism="Sus scrofa"
   /strain="University of Nebraska, Lincoln Swine Selection
```

lines-
/db_xref="taxon:9823"
/clone="UNL-P-FN-cx-a-05-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pP73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI. The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares. Genome Research 6: 791-806
1996.

	TAG_SEQ=None found"							
BASE COUNT	244	a	149	c	161	g	176	t
ORIGIN								

Query Match	2.3%;	Score 69;	DB 11;	Length 730;
Best Local Similarity	100.0%;	Pred. No. 6.1e-08;		
Matches	69;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Dy 2374 aaagaatggygaatgaaagatgaaaatgcacctaaccctccagaacccaagaaagtta 2435
|||||
Dd 649 AAGAAGTGGGGAATGAAGAATGAACCTAACCCCTCCAGAACCCAGAAATTAA 708

QY	2434	ataaaaagc	2442
Db	709	ATAAAAAGC	717

[illegible]

ORGANISM	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
AUTHORS	1 (bases 1 to 468)
TITLE	Caetano, A.R., Johnson, R.K. and Pomp, D.
JOURNAL	Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
COMMENT	Unpublished (2001)
	Contact: Pomp, D

Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomeunl@unl.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail.
Seq. primer: M13 -29
POLYA=yes.

FEATURES
Source

```
/organism="Sus scrofa"  
/strain="University of Nebraska, Lincoln Swine Selection  
Lines"  
/db_xref="taxon:9823"  
/clone="UNL-P-FN-bo-a-05-0-0-UNL"  
/clone_lib="UNL-P-FN"
```

`/dev_stage="Adult"`
`/lab_host="DH10B (Life Technologies)"`
`/note=Vector: pT73D-pac (Pharmacia) with a modified`
`polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-p-FN`
`library is a normalized library representing porcine`
`ovarian follicles, ranging between 2.0 to 10.0 mm in`
`diameter, collected during 7 days of the follicular phase`
`of the pig estrous cycle. This library was derived from`
`the library UNL-p-F2. The tag is a string of 5-6`
`nucleotides present between the Not I site and the`
`oligo-AT track. The library was constructed as described`
`by Bonaldo, Lennon and Soares, Genome Research 6: 791-806`
`1996.`

	TAG_SEQ=None found"			
BASE COUNT	127 a	98 c	97 g	145 t
ORIGIN				1 others

Query Match	2.3%;	Score 68,	DB 11;	Length 468;
Best Local Similarity	100.0%;	Pred. No.	1.6e-07;	

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2884 gactctggccccgagcgcttggaataaataactctgtctgttgcacataaaaaa 2943
 |||||||
 Db 68 GACTGTGGCCCCGAGCGCTTGGAATAAATCTCTCTGTTCATCAAAAAA 9
 |||||||

Oy 2944 aaaaaaa 2951
 |||||||
 Db 8 AAAAAA 1

RESULT 10
 BF703671/c 485 bp mRNA EST 22-DEC-2000
 LOCUS MI-P-E4-abp-c-04-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
 DEFINITION MI-P-E4-abp-c-04-1-UM 3', mRNA sequence.
 ACCESSION BF703671
 VERSION BF703671.1 GI:11989079
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE 1 (bases 1 to 485)
 JOURNAL Normalization and subtraction: two approaches to facilitate gene
 MEDLINE discovery
 COMMENT Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Tugale CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ctugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized embryo at gestational day 14 library cDNA library
 preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
 Research Center, Department of Animal Science, University of
 Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source Location/Qualifiers

1..485
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-E4-abp-c-04-1-UM"
 /clone_1ib="MI-P-E4"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-E4
 library is derived from embryo at gestational day 14. For
 a detailed description of the library from which this
 clone was derived, please visit our web site at
 http://piglet.genome.iastate.edu/
 TAG_LIB=MI-P-E4
 TAG_TISSUE=embryo at gestational day 14
 TAG_SEQ=ATCATC"
 BASE COUNT 142 a 114 c 109 g 119 t 1 others
 ORIGIN

Query Match 2.2%; Score 64; DB 11; Length 485;
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2819 aaccataaagctgtccgacttcgcactcggggccgagctctaccctcgtcgtg 2878
 |||||||
 Db 131 AACCCATTAAGCTGTCCGATTCGCCGACCTCGGGGCGCACTCTCAACCCCTCGTGG 72
 |||||||

Oy 2879 tgta 2882
 |||||
 Db 71 TGTA 68

RESULT 11
 BF703805/c 554 bp mRNA EST 22-DEC-2000
 LOCUS MI-P-O2-abs-a-11-1-UM.s1 MI-P-O2 Sus scrofa cDNA clone
 DEFINITION MI-P-O2-abs-a-11-1-UM 3', mRNA sequence.
 ACCESSION BF703805
 VERSION BF703805.1 GI:11989213
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE 1 (bases 1 to 554)
 JOURNAL Normalization and subtraction: two approaches to facilitate gene
 MEDLINE discovery
 COMMENT Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Tugale CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ctugale@iastate.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized ovary at estrus day 5 library cDNA library
 preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
 Research Center, Department of Animal Science, University of
 Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source Location/Qualifiers

1..554
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-O2-abs-a-11-1-UM"
 /clone_1ib="MI-P-O2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O2
 library is derived from ovary at estrus day 5. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 http://piglet.genome.iastate.edu/
 TAG_LIB=MI-P-O2
 TAG_TISSUE=ovary at estrus day 5
 TAG_SEQ=ATCATC"
 BASE COUNT 158 a 128 c 122 g 146 t
 ORIGIN

Query Match 2.2%; Score 64; DB 11; Length 554;
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2819 aaccataaagctgtccgacttcgcactcggggccgagctctaccctcgtcgtg 2878
 |||||||

REFERENCE 1 (bases 1 to 300)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail.

Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..300
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-ck-g-02-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
 TAG_SEQ=None found"

BASE COUNT 69 a 62 c 71 g 97 t 1 others
 ORIGIN

Query Match 1.7%; Score 49; DB 11; Length 300;
 Best Local Similarity 100.0%; Pred. NO. 0.0094;
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2884 tgactgtggcccgagcgttgaaataaacctctgtgttgcga 2932
 ||||||||||||||||||||||||||||||||||||||||
 Db 66 GACTGTGGCCCGCAGCCGCTTGGAATAAATCCTTCTTGTTCGA 18

RESULT 15
 B1182742/c 494 bp mRNA EST 10-JUL-2001
 LOCUS B1182742
 DEFINITION UNL-P-FN-bo-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-bo-e-05-0-UNL 3', mRNA sequence.
 ACCESSION B1182742
 VERSION B1182742.1 GI:14657151
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 494)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science

University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..494
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-bo-e-05-0-UNL"
 /clone_lib="UNL-P-FN"
 /dev_stage="ADULT"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
 TAG_LIB=UNL-P-FN
 TAG_TRISUP=porcine ovarian follicles
 TAG_SEQ=CACT"

BASE COUNT 118 a 117 c 125 g 134 t
 ORIGIN

Query Match 1.6%; Score 48; DB 11; Length 494;
 Best Local Similarity 100.0%; Pred. NO. 0.01;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2889 tggcccgagcgcgttgaaataaacctctgtgttgcacaa 2936
 ||||||||||||||||||||||||||||||||||||||||
 Db 64 TGGGCCCGCAGCCGCTTGGAATAAATCCTTCTTGTTCACAA 17

Search completed: February 24, 2002, 04:28:52
 Job time: 24078 sec

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@email.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tldb/Bac_ends/mouse/bac_end_intro.html
Plate: 149 row: D column: 1
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostom

Ddb QY 823 gtcaggaacgtgacaatctgacttggagacagaataaactgaatagcccccatgca 882
|| || | ||| ||| | ||| | ||| | ||| | ||| | ||| |
361 GTTAAAGATACGGCAAGCTTACTCTAGGTGAGTCGTGAATCATTACATCCCCCATGCC 420

[illegible]

RESULT	3
LOCUS	AZ329067/c
DEFINITION	AZ329067 673 bp DNA GSS 29-SEP-2000
ACCESSION	U0053G10.F Mouse 10kb plasmid U0053G10 library Mus musculus genomic clone U0053G10M0053G10 F, DNA sequence.
VERSION	AZ329067
KEYWORDS	AZ329067.1 GI:10389414
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 673)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly ,M., Rose,M., Rose,R., Stokes,R., Tinsley,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunne@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0053 row: G column: 10 Seq primer: CGTTGTAAACGACGGCCAGT Class: plasmid ends High quality sequence stop: 673. Location/Qualifiers 1..673
FEATURES	
Source	

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0053G10"
/clone_lib="Mouse 10kb plasmid U06C1M Library"

```

/sex="Male"
 /lab host="E. Coli strain XL10-Gold, T1-resistant, F"
 /note="Vector: pWD42hv: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g11473211419b|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match	11.2%	Score 373.2	DB 13	Length 673
Best Local Similarity	78.1%	Pred. No. 3.5e-64		
Matches 520	Conservative	0	Mismatches 123	Indels 23
			Gaps	5
QY 1900	ctaaggaagcaccaccagcgctactacttggagatggacttacttgaaglaaacgcgct	1959		
Db 672	cttcacggatggcttgctctgaacccacttctggcaggtagatgttaccgaacccacgct	613		
QY 1960	aaatccggaacaaatatctatgttlttggaaacacttccagatgggtaagagct	2019		
Db 612	acatccggcaacaaagacttattttaggtttatagatcccttttcaaggctgataagct	553		
QY 2020	tatctactaagaagaagacttcaacccgtygttggcttaagaataacttgaagaaatttt	2079		
Db 552	ttttctaccacgacactgagac-----ggccaatatgatcctaacgcaaatcttc	505		
QY 2080	ccaagatttggaaatcacttaagtaaggtcaagaacatggtccagcttctgttgcag	2139		
Db 504	ccgagctttgtatattaccacaggttaacaggtccaaataatggactctcttgttgcacag	445		
QY 2140	gtaagtcagagacttgcacagatatttgggatttgcatttggaaactgacttgttgatacaga	2199		
Db 444	gtatagtcagagacttgcccaagatatttagagatttggtttagaagttgtacattgtgcatacag	385		
QY 2200	ccccaaagctcagagacaggtagaagatgatagataccaattaaagaagacccttaccaa	2259		
Db 384	ccccaaagctcagagacaggtagaagatgataccaattaaagaagacccttaccaa	326		
QY 2260	ttgacacagagagacttgcatttaattgatttggatgtgcctccctgcacttgcatttga	2319		
Db 325	tttgcacgtcagacactgagactaatgatataatgatatgcttacccttgggtccttcag	266		
QY 2320	gtgagagaacaccccttggacagtttgggtctgaacccctatgaattgtctctac-----	2375		
Db 265	gttagaacaacatccctgaacagtttggactgaacccctatgaattgtctctctac-----	206		
QY 2376	acccccccgcttggcagaatctgccttgcataatgctgcatagttgcattgtcttccagcc	2435		
Db 205	ggggcctccactgtaaaaaatagcctctgtacatagtgctgcacatcgtctgttccagcc	146		
QY 2436	tttgttctctaagctcaagcgctcgatgggttgaagcgaga-----gcttgaagcagc	2491		
Db 145	tttgttctctaagctcaagcgctcgatgggttgaagcagcagcagcagcagcagcagc	86		
QY 2492	ttccggagagcctactcagag--gagacttgcagattccacatcgcttcccaatttga	2549		
Db 85	ttccgaagagcctactcagagagacagcactgcaaccccccacatctcttacaagtttaaga	26		

QY 2550 ttcagt 2555
 Db 25 TTCGGT 20

RESULT 4
 LOCUS B1107577 799 bp mRNA EST 26-JUN-2001
 DEFINITION 602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
 mRNA sequence.
 ACCESSION B1107577
 VERSION B1107577.1 GI:14558470
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 799)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM1107 row: n column: 05
 High quality sequence stop: 743.
 Location/Qualifiers
 1. 799

FEATURES
 source
 /organism="Mus musculus"
 /strain="C57BL/6J (f1)"
 /db_xref="taxon:10090"
 /clone_1lb="NCI_CGAP_Lu29"
 /issue_type="spontaneous tumor, metastatic to mammary."
 /stem_cell_origin="Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 211 a 215 c 218 g 154 t 1 others

ORIGIN

Query Match 10.5%; Score 349; DB 11; Length 799;
 Best Local Similarity 70.6%; Pred. No. 2.2e-59;
 Matches 537; Conservative 0; Mismatches 215; Indels 9; Gaps 5;

QY 729 gaagagacgtgtctactgtcaagaagcttgatctgtgacgaagtggtgacct 788
 Db 19 GCGGGCCCGGTGCTACTGTATCAAGAACTGGACGCTGTGGCCAGTGGCCCTC 78

QY 789 atgtctgaagctatcgacgtgtgtgcaactgtgtcaagaagcgtgacaattgactt 848
 Db 79 CTGCTCGCGAGCATGACGACGCGCTGTGTAGTACAGATGCTGCAAACTGACTAT 138

QY 849 gggagacgaataactgtatagcccccatgcatgtggaagaacatcgctggagccccc 908
 Db 139 GGGCCGGAATGTACTATATGTGGCCCAACATCTCTTGAGACATCATCAGCAACACC 198

QY 909 agaccgatgatgacccaagcccgatgacccactataaagcctgcttcaagaag 968
 Db 199 GGACGCGCTGATGACCAAGCCCGGATGACGACTACAGACCTATTGCTGACAGACG 258

QY 969 ggttaacttcgtccacacagccgctctcaaaccttccactctctgctgaagaactga 1028
 Db 259 AGTAGATTTCACCCCGCAGCATTCATCAACCCCGCTCTTACTACTCT---GAGGCTGA 315

QY 1029 tgaaccagttactcatatgtattgtccatcaactatgtattgaggagacttgggtccgcaagga 1088
 Db 316 CGAGGCCCTGTCACATTAAGTGTGAGAAATATCGGACAGAGATGGAATCCGGCCAGA 375

QY 1089 ccttacagacataccgttgaacttgagaagtgctcaactctgttactgaaggaagcctca 1148
 Db 376 CTTACAGACCAACCTTGGCCAGGCGGATG---ACTGTGTTACGACGGAAGCGAGCTT 432

QY 1149 tctgtgtgaaggaaggaaggaatgagctgtggcgagctgtgtgagagagccgc-acgactc 1207
 Db 433 TGTGTGAGAGGTAAAGCGGAAGCTGGGGCTGCAGTAGAGGAGAAACGGCTGATATAT 492

QY 1208 gggccagcagcctgcccgggaaggaacttcgcgcaaaaagctgagctcatgcccctaagc 1267
 Db 493 GGGCCAGCACTGTGCGGAGGATCATGACCTTAAACCGGAATCAATCGCATTAATTC 552

QY 1268 aagcttgcgctgcccgaaggaaggaatccataacattatacagacacaggtatgct 1327
 Db 553 AAGCTTAAGGCTGGGCAAGGAAGGCTCTTAATGTCTATACGACACCTGTACGCTT 612

QY 1328 ttgcgactgacacgttacacggtggtccatataaacaaggggtgttactcaagcag 1387
 Db 613 TTGCCAGGCTCATGTTTCACGAGCAATATACGACACCGTGGAGCTGACGTGACG 672

QY 1388 gggaggaataaagaaca-aagaggaattcttaagctcttagaagcct-tacattgctc 1445
 Db 673 GCAAAAGATATCAAAATATACGAAAGAAATTCACGCTTAATTAAGACCTGTTCATTGCGC 732

QY 1446 aaaaagctagctatatacactgtctctgtgacatcaagaag 1486
 Db 733 CGGTAGGTGGCAATTATTCATTGCCAGACCAAGCAAGG 773

RESULT 5
 A2337037 705 bp DNA GSS 29-SEP-2000
 LOCUS 1M0067L03R Mouse 10kb plasmid U06C1M library Mus musculus genomic
 DEFINITION clone U06C1M0067L03 R, DNA sequence.
 ACCESSION A2337037
 VERSION A2337037.1 GI:10406938
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 705)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0067 row: 1 column: 03
 Seq primer: CACACAGGAATACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 705.
 Location/Qualifiers
 1. 705

FEATURES
 source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C1M0067L03"

[illegible]

FEATURES		Location/Qualifiers	
source		1..789	
		/organism="Mus musculus"	
		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/clone="RPci-24-155A1"	
		/clone_1lib="RPci-24"	
		/sex="Male"	
		/cell_type="Spleen/Brain"	
		/note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI; RPci-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."	
BASE COUNT	230 a 200 c 195 g 164 t		
ORIGIN			
Query Match	10.2%; Score 340.2; DB 13; Length 789;		
Best Local Similarity	.65.5%; Pred. No. 1.2e-57;		
Matches 514; Conservative	0; Mismatches 268; Indels 3; Gaps 1.		
QY	789 atgctgaagcgtatcgccgctgtgacactgtgtcgaagcgcctgcataatgacctt	848	
Db	4 atgtctgaagcgtatgtgcagcagtgacccctgttgatttaagatgctgcacaaattgcacat	63	
QY	849 gggacagaaataactgttaataagccccccatgcatgtgagaaacatcgcttcgacagcccc	908	
Db	64 gggacagaaagggagcagtggtgtgccccctcatgcttgaagaaatgctgcacagccacc	123	
QY	909 aagaccgtatgatacgaacagccgcgatataccacatataaacctgccttcacagagag	968	
Db	124 tgcacagatgcatgacaaatagcccgatgccacacatcagacccctgcttaaatgattg	183	
QY	969 ggtcaccttcgcttcacacagccgcctcacaaccttcacaccttccttcgcttgaagagactga	1028	
Db	184 tgtataccttttgcccccctgcctccttcaaacacactgaccccttcccttacaacaaatga	243	
QY	1029 tgaaccagtgactcatgtattgcacatcatattgattgaagagactggtgtccgaaaga	1088	
Db	244 ttccgtccacagtcacatcattatgtatggaacatcttcctcaagaaatctggaccagaaatga	303	
QY	1089 ccttacagacatataccgcctgcactggaagtgcttaacctgttcacttgacggaagagcta	1148	
Db	304 cctgacatgacccaacccctggcctg---agctccacagtttgatgacacgacgcagcagattt	360	
QY	1149 tctgtgtgaagtgtaagagatgctgtggcgagctgtgtgacgagaccgcgacacatctg	1208	
Db	361 cctgtatgaagggaagaaacgaagcgtgcagctgcggtgtgatatgggaaaggaatgatttg	420	
QY	1209 ggcgaagcagcctgcgcggagaagaacttaagtcgcaaaagctgaagctcattcattgcctcaagca	1268	
Db	421 ggcgaagccttttgccttgagagacgtgcgcacaaagccttaacttataaggccttatttca	480	
QY	1269 agcttgcgctctggccgaaggaatccataaacattatatacggacagcagatgtacctt	1328	
Db	481 agcccttcgagagagcttaaggtatgatttaacattcagactacacacccctatgcttt	540	
QY	1329 tgcgactgcacaagtgacaggggcacatcataaacaagggggtgtccttaccctcaagag	1388	
Db	541 tgcctacagcacacatccatgaggccatcttaccagcagcagcctatttgacttcgctgg	600	
QY	1389 gagggaataaagacaagaaggaatcttaagaccttgaagccttaacttgcacaa	1448	
Db	601 taaagcattttaaatacaagaaagaaatttctggccctgttggaacccatcactgcacctaa	660	
QY	1449 aagcgtactatatacactgtccttgacatcagaagaacaaatcatatctagag	1508	
Db	661 gaaggtgacacatcattccacgtgccccgggacacaaagaggaagacacttgggtggccacgg	720	

QY 1509 gaaccagatggtccggttgcgaagcagcagccagcgttgaaccttgcctat 1568
DB 721 CCAACGAGTGCAGACTCAGTACGAAACAGTGTCTCAAGGGCCATGATCTTACTGA 780
QY 1569 atag 1573
DB 781 AAAAG 785

RESULT 8
AZ715272 755 bp DNA GSS 24-JAN-2001
LOCUS RPCI-24-154N5.TV RPCI-24 Mus musculus genomic clone RPCI-24-154N5,
DEFINITION DNA sequence.
ACCESSION AZ715272
VERSION AZ715272.1 GI:12451823
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 755)
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-154N5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 154 row: N column: 5
Seq primer: SP6
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..755
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-154N5"
/clone_1lb="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARABAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 202 a 194 c 190 g 169 t
ORIGIN

Query Match 10.2%; Score 338.8; DB 13; Length 755;
Best Local Similarity 68.1%; Pred. No. 2,3e-57;
Matches 501; Conservative 0; Mismatches 222; Indels 3; Gaps 2;

QY 49 cccactagcaaccactttgtccttcgaatgagagatccaggtacggagaagcaacg 108
DB 23 CCCAACAGCCAGCCCTGTGCTTGCATGATGAGACTCCGAGAGTGACACGCGGA 82
QY 109 cagctcacttgagcccgatcccaagaggttcagaactcccgacacatcttgaagaa 168
DB 83 CAGCTCACATGAGCAGAGGCTCCTCAGGATTCAGAAAGACTGCGCCACTTTGTCGATGAA 142

QY 169 gccctacacagggagactggccaaactcagatccaaacacctcaggtgacctctccag 228
DB 143 GCCCTACACGAGATGATCTTCTTTCAGAGCAATAAACCCACAGGTATCTTCTGCAA 202
QY 229 taagtgtgagactgtcttcgtgagagccacaaacagactgtctagaagtacgaag 288
DB 203 TATGTAGATGACCTGCTCTCAGTGCAGAAACAGTGAAGCTGTGAATTTGGACCCAA 262
QY 289 gcaactgtcgtgaattgtcgtacgtcaggtcacaagagctctgtcctaagaagccagatt 348
DB 263 AACCTCTGTGGGCGAGTTAGTAAAGCTGGGATAGCGGCTCTGTCTTAAAGGCTCAGTTA 322
QY 349 tgcagagagaggttaacatacttgggtacagtttcgagggcgagcagatggtcagc 408
DB 323 TGCCAGATTAAGATGACCTAGATGATGATGCTTGTAGAGATGACACAGGTGGCTCACA 382
QY 409 gaggcagagagaagaactgtatccagataccggcccaacacagcaacaagaatgaga 468
DB 383 AAAACCGAAGAACACCTGTATGACAGATCCGACCCCAACCTGCTGCGCAGGTAAAG 442
QY 469 gaattttgggagcagctggaatttgcagactgtgacccggggttgcagacttagca 528
DB 443 GAGTCTCTGGGGACCGCCGGGTTTTCAGAGCTGTGATTCGCGATTGGCCACACTAGCA 502
QY 529 gccccactcaccgctaaccaaaagaaaggggagttctcctggctcctgagcaccag 588
DB 503 GCTCCCTTGTATCCACTAACCAAGAGAAAGGGGATTCACCT-GGATGAGAACATCAG 561
QY 589 aagcatttgatgtctcaaaaagccctgctgagcagctgtctgcccctctgac 648
DB 562 CTGACCTTTGAAACCTCTCAAAAAGGCACTGTCAGGCTCCGCGATTTGGCTCCAGAT 621
QY 649 gtaactaaccttacccttacccttgaatgagcgaaggaatgagccagagattca 708
DB 622 TTAACAAACCTTTCAACCTATACATTGATGAGAAGAAATGAGATGGCAGAGGCTT 681
QY 709 acccaaaccttagaacatagagagacgtgtgtcctacactgcaagaagcttactct 768
DB 682 ACCCAGGTTTGGGACATGAGAGCGCCGCTAGCTACTTATAAGAAA-CTGACGCT 739
QY 769 gttagcagtggttggc 784
DB 740 GTGGCCAGTGGATGGC 755

RESULT 9
BH083773 653 bp DNA GSS 18-JUL-2001
LOCUS RPCI-24-293G13.TV RPCI-24 Mus musculus genomic clone RPCI-24-293G13
DEFINITION DNA sequence.
ACCESSION BH083773
VERSION BH083773.1 GI:14903370
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 653)
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-293G13.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong

QY 520 accttagcagccaccctctaccgcttaacccaagaaggggattctctctgggctcct 579
 Db 68 ACCCTAGCAGACACCTCTATATCCTTTGACATAGAGAGGGGCTCTTGCAGTGGAAAGAA 127
 QY 580 gagacacagaagagctttagtctatcaaaaagccctgtctgagcagacctgtctggcc 639
 Db 128 GAACACACAGAGAGCTTTTGGAGGCTATTAAGTGGCTCTATAGACTGGCCCCGGGCTAGCA 187
 QY 640 ctcccttagcagtaacaaaccttacccttattgtgtgagcgttaagagagtagccga 659
 Db 188 TTACACAGCTTGATGACCTTTCTGCTCTATATGTGAGCAGAGAGCGGGTGTAGCCAGG 247
 QY 700 ggaatttaacccaacccatgagacccatgagagagaccttgcctaccctgtcaaaag 759
 Db 248 GGAATGTTGACACAGAGAGCTGGACCTGTGAGAGAGACCTGTAGCTTATTGTCAAAAAA 307
 QY 760 cttagccttagcagcagtggttggccggtatgtctgaaagcctatcgagcgtgtggcata 819
 Db 308 TTAAATCTCTGTGCTAGTGGATGGCCACATGCTGAAAGCTATWGACAGCAGTAGCCCTG 367
 QY 820 ctgtgcaagaagcgttgacaaatgtgacttggagacagataataactgttaagcccccac 879
 Db 368 TTGATCAAAAGATGCTGACAAATTTGACAAATGGACACAGGTGACTGTTGTGGCCCTCAT 427
 QY 880 gcatltagagacacatcgttgcgacgcccccaagccagatgagtagacacagcccgatgac 939
 Db 428 GCTTAGAAATGATGCTGCGGACGACCTGACAGATGATGACAAATGCCCGGAAATGACA 487
 QY 940 caactcaaaagcgtctctacacagagaggttcaacttgcctccaccagcgcctctcaac 999
 Db 488 CACTATCAGAGAGCTGCTCTCTAAATGAGGGGTAACTGTTGGGCCCTGCATCCTCAAC 547
 QY 1000 cctgaccactctctgctcgtgaagagactgtgaaacagatgactatgttgcacacta 1059
 Db 548 CCAGTACCTCTTCCCTCTTAACAAATGATTCGTCGCCAGTACATCATGATGAGACATC 607
 QY 1060 ttgataggagagactgggttcgcgaagaccttaacagacatcacctgagactgag 1114
 Db 608 CTGTGTAAGAAATGAGGACCAAGAGTACTGACTGACCAACCTGCGCTGGAG 662

RESULT 11
 A2897994/c
 LOCUS RPT-24-173A19.TV RPT-24 Mus musculus genomic clone RPT-24-173A19
 DEFINITION
 , DNA sequence.
 ACCESSION
 A2897994
 VERSION
 A2897994.1 GI:13216939
 KEYWORDS
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 1 (bases 1 to 689)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Ahirel,B., Levins,M.,
 Tseay,G., Geer,K., Krol,M., Shartsbeyn,A., Gebregorjais,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPT-24
 Unpublished (1999)
 Other_GSSs: RPT-24-173A19.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPT-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 173 Row: A Column: 19

Seq primer: 77
 Class: BAC ends.
 Location/Qualifiers
 1..689
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPT-24-173A19"
 /clone_lib="RPT-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site.1: BamHI; Site.2: BamHI.
 RPT-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 168 a 174 c 170 g 177 t
 ORIGIN

Query Match 10.0%; Score 33.2; DB 13; Length 689;
 Best Local Similarity 68.9%; Pred. No. 3; le-56;
 Matches 474; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

QY 1933 gtggaactcaactgaggaagccgcttaaatcaagaacaataatctatgttlttga 1992
 Db 688 GTGATTTTACTGAAGTTAAACCTGGAATGATGTAAAGATCTGTAGTATTGTA 629
 QY 1993 gacacatttcaagatgggtagaggttaccctactaagaagaagacttcaacggtgt 2052
 Db 628 GACACCTTTTAAAGATGGGTCGAGGCAATTTCCACTAAACAGAGACTGCCAGATTGG 569
 QY 2053 gctaaagaanaactgaggaagatttcccaagaatttgaatacctaaggttaaggtgca 2112
 Db 568 GCCAAGAGATGCTCTTGAAGAAATCCGCCAAGATTTGGAATCCCTAAGTAATCGGGTCC 509
 QY 2113 gacaatgtgccagcttctgtccaggtgaagtcagaggtgccaagatatgggatt 2172
 Db 508 GACAATGAGACACACCTTTGTCGCCAGTAAAGTACAGGCTGACCTGATGGGCTATC 449
 QY 2173 gattggaactgattgtgtatagaagcccaagctcaagacaggtagaagatgaaat 2232
 Db 448 GATTGAAATTTACCTGTGTGTACCGCCCTCAAAAGCTCAGGACAGATGAGAGATGAAT 389
 QY 2233 agaaccattaaagagaccttaccaaatltgacacagagactggtcaatgattgattg 2292
 Db 388 AGGACCTTAAAGAGACTTGACTAATTTAGCCATTGAGACCGCGGGAAGACTGGGTG 329
 QY 2293 gctctcctgaccttgtgtctttaggtgtgagaaacccctgtgacaggttgggtgacc 2352
 Db 328 GCTCTCTCCCTCTTGTGGCTTCCGAGCCCGGACACTCTGAGCGTTTGGGCTCAGT 269
 QY 2353 cccatgaattgctctacagggagacccccccgcttgcagaatgtg-----ccttgc 2406
 Db 268 CTTTAAAGTCTGTGTGAGAGACTCCGCCCTTATGAGAGGTGTAGAACATTTAGT 209
 QY 2407 catagtgtgattgtgtgtcttccagccttgttctgtgctcaagggcctcgattg 2466
 Db 208 TCCGACTCTGACCTGTCTTACCTCCTCTTGTATTCAAAAAGGCGCTTGAAGAGT 149
 QY 2467 gtgagcagcagcaggttgaagcagctcgcgagcctcaagcaggtgagagactt 2526
 Db 148 ATTGAACCCGATTTTGGACCACTGAACACAGCTATACCCAGGAGACCGACCGAGTA 89
 QY 2527 ccacatgcttcaagttgagattcagctatgtttagagccacgcttgagaagaaactc 2586
 Db 88 CCCGACGGGCTTCCGAGTTGGAGACAAAGTCTTGTCAAGCGCGCATTAAGACCGGAGCTT 29
 QY 2587 gagactggttgaaggtacttactcg 2614
 Db 28 GAGCCACGCTGGAAGGACCTATTGG 1

RESULT 12
 LOCUS A2717731 817 bp DNA GSS 24-JAN-2001
 DEFINITION RPI-24-15511.TV RPI-24 Mus musculus genomic clone RPI-24-15511,
 DNA sequence.
 ACCESSION A2717731
 VERSION A2717731.1 GI:12456936
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 817)
 Zhao, S., Niemann, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tseng, G., Geel, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 Other GSSs: RPI-24-15511.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 155 row: 1 column: 1
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1. 817
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-15511"
 /clone_1lb="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pFARBAC1; Site.1: BamHI; Site.2: BamHI;
 RPI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pFARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 BASE COUNT 243 a 209 c 204 g 161 t
 ORIGIN

Query Match 10.0%; Score 331.4; DB 13; Length 817;
 Best Local Similarity 65.5%; Pred. No. 6.7e-56;
 Matches 501; Conservative 0; Mismatches 261; Indels 3; Gaps 1;

QY 809 ctgtgacatgctgcaagagcgtgcaaatgacttgggaagaataactgttaa 868
 1 CAGTACCCCTGTGATTAAAGATGCTGACAAATGACAAATGGACACAGGCTGTG 60
 QY 869 tagcccccacatgctgcaagacatgctgcaagccccacacgatgatacag 928
 1 TGGCCCTCATGCTTAGAAATATCGCGCTGACACCATGACAGATGATGACAAATG 120
 DB 61 TGGCCCTCATGCTTAGAAATATCGCGCTGACACCATGACAGATGATGACAAATG 120
 QY 929 cccgatgacccatcataaagcgtctctacagagaggtcacttgcgtccacag 988
 121 CCCGATGACCATATAGACCTGCTGCTAAATAGCAGCTGTAACTTTGGCCCTG 180
 QY 989 ccgcctcaacccctcactctctgctgagagagactgatacagacagtgactat 1048
 181 CCATCTCAACCCAGCTACCTTCTCCCTTAACAAATGATTCGTCACCATGATCAT 240
 QY 1049 gccatcaactatgattgagagagactgggtccgcaagacattacagatacgcgtga 1108

DB 241 GTATGACATCCCTCGTGAAGAACTGGACCAAGTACCTGATGACCAACTCTGC 300
 QY 1109 ctgagagaagtgtctaactgttactgacggaagcaagctatgtgtgaagtaagga 1168
 DB 301 CTGG--AGCTCCAGTGTGTACACGAGCGAGAGCTTCTGATAGAGGAAACGAA 357
 QY 1169 ttgctgggagcgtgtgtgagcagccgcacgcttcggccagcagctgcgggaag 1228
 DB 358 AGCTGAGCTGCTGCTGTGTGATGATGGAATAATTTGGCAACGCTTTGCCGGAAG 417
 QY 1229 gaacttgaagcacaagagctgagctatgcccctcagcaagcttgcgtgcggcaag 1288
 DB 418 GAACGTGCGGACCAAAAGGCTGAACCTATAGCGCTTTTAAACCCCTCGAGAGCTTAAG 477
 QY 1289 ggaatccataaactattataaggaagcagatgacttgcgtgcagtcacagtaacg 1348
 DB 478 GTAAGATCAATTAACATCTACACGACGCGCTATGCTTTCTACAGCACATCATG 537
 QY 1349 gggcatctataaacaaggggtgtgttactctgacgagggggaataaagaacaag 1408
 DB 538 GGGCCATCTACAGCGCAAGGCGTATTACCTCGCTGTGAAGACATTAATAAACAAG 597
 QY 1409 aggaattctaagcctattagaagccttacttgcacaaagagctagctattacact 1468
 DB 598 AAGAAATCTGCGCTGTGGAAGCCATACATGACACTTAAGAAAGTTAGCATCATCACT 657
 QY 1469 gtcctgagatagaagaagcacaagatctcatatctcaagaggaacagatgctgcgg 1528
 DB 658 GCCCGGCGCACAAAGAGGAGAAAGCTTGTTGGCCAGGCGCAACCGAATGGCAGACTCAG 717
 QY 1529 ttgccaagcagcagccagcgtgttaacctctgcctataatag 1573
 DB 718 TAGCAAAACAAGTGTCTCAAGGGCGCATGCTTACTGAAAAG 762

RESULT 13
 A2769626 692 bp DNA GSS 16-FEB-2001
 LOCUS 1M0570J04R Mouse 10kb plasmid U00C1M library Mus musculus genomic
 DEFINITION clone U00C1M0570J04 R, DNA sequence.
 ACCESSION A2769626
 VERSION A2769626.1 GI:12889953
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 692)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0570 row: 3 column: 04
 Seq primer: CACACAGAAACAGCTATGAC
 Class: plasmid ends
 High quality sequence stop: 692.
 Location/Qualifiers
 1. 692
 /organism="Mus musculus"
 /strain="C57BL/6J"

TITLE
 JOURNAL
 COMMENT

FEATURES
 source


```

Db      548 ATCGGGTCGACAAATGAGACCACTTTGTGGCCAGGTAAGTCAGGGCTTGACCACTCAG 489
Qy      2164 ttggggatttgatggaacatcgtatgcatatagaccccaagctcaggacagtagag 2233
Db      488 TTGGGCAATGATTTGAAATTAACACTGTCTTACCGCCCTCAAGCTCAGACAGGTAGAG 429
Qy      2224 aggaatagatgaacacctaagaagacccctacacaaattgaccacagagactcgacataat 2283
Db      428 AGGATGATGATGAGACTTAAAGAGACCTTGAACCTAAATTAAGCCATTGAGACCGGGGAAA 369
Qy      2284 gattgatgctctcctcgtcccttctgtctttagagtgaggaacaccccgagacagttt 2343
Db      368 GACTGGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
Qy      2344 gggctgaccccttgaatttctctcagggagaccccccccgcttgacagaatg----- 2398
Db      308 GGGCTCACTCTTTTGAAGTCTGTATGAGAGACCTCCCTCAATGAAGCTGGTGA 249
Qy      2399 -ccttgacatagtgctgatgtgtcttccagccttctctcctagcctcagagcg 2457
Db      248 ACATTGGTTCCGACTGTGACCTGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 189
Qy      2458 ctgagtggttgagcagcagcgttggaagcagctccggagagcactactcaggagagac 2517
Db      188 CTAGAACTGATGAGACCAAAATTTGGACCACTGAAGCAGCCTATACCCAGGAGAC 129
Qy      2518 ttgcaagttccacatgcttccaaattgagatcagctatgtatgagccacagctgca 2577
Db      128 ACCGCACTACCCACGGGTTCCGAGTTGGAGACAAAGCTGTGGTCAAGCGCATGCAACC 69
Qy      2578 ggaacactcagagactcgttggaagagacactatcgt 2615
Db      68 GCGAGCCTCGAGCCACGCTGAGAGGAGACCTATTGTGT 31

```

RESULT 15

AZ432720

LOCUS

AZ432720 610 bp DNA GSS 03-OCT-2000
 1M0218E17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0218E17 F, DNA sequence.

DEFINITION

AZ432720
 AZ432720.1 GI:10556733

ACCESSION

AZ432720
 GSS.

KEYWORDS

house mouse.
 Mus musculus

SOURCE

house mouse.
 Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 610)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, J., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.
 and Wright, D., Weiss, R.

AUTHORS

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

TITLE

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

JOURNAL

Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

COMMENT

Insert Length: 10000 Std Error: 0.00
 Plate: 0218 row: E column: 17
 Seq primer: CGTTGTAAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 610.
 Location/Qualifiers
 1. 610

FEATURES

source
 /organism="Mus musculus"
 /strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="UUGC1M0218E17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/notes/). The DNA was hydrolytically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (914732114191AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was ligated to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 170 a 148 c 133 g 159 t
ORIGIN

```

```

Query Match 9.9% Score 327.6; DB 13; Length 610;
Best Local Similarity 82.6%; Pred. No. 4.1e-55;
Matches 436; Conservative 0; Mismatches 84; Indels 8; Gaps 5;

```

```

Qy 2110 tcaagaatggtccagcttctgttccagatgtaagagacgtgccaagattggg 2169
Db 1 TCACATTAACAGACCGCTTTGTTGTCCTGAGTAAATCAAGGACGCGCAATATTGGGG 60
Qy 2170 attgattggaacatgcatgtgcatagacagcccaagctcagagagtagaagatg 2229
Db 61 ATTGATGATGAGATTAATTTGATGATGATGATGATGATGATGATGATGATGATG 120
Qy 2230 aatgaaccatlaaagagaccccttaacaaattgacacagagagcattgaattgg 2289
Db 121 AACGAAACCAAT-AAAGAGACCCCTTACCAAAATGACAGACAGACTGCGCTAATGATGG 179
Qy 2290 atggctcctccgttctgttctttagagtgagagaccccttgacagcttggctg 2349
Db 180 ATAGCTCTCCACCTCTTGTGCTCTTCAAGGTTAAACACCCCTGACAGTTGGACTG 239
Qy 2350 acccctatgaattgtctac--ggggagaccccccgttgagagaattgtcctggac 2407
Db 240 A-CCTTATTAATCACTCTACGGGGGGGGGAGCTCCACTGATGATGATGATGATG 298
Qy 2408 atagtgctgtagtgccttccagccttcttctcctagctcagagcgtcagtgag 2467
Db 299 ATAGTCTGACATGCTGCTTCCACGCGCTTCTTCTGAGCTCGAGTCTCAATGAGG 358
Qy 2468 tgaagcagcagcgttgaaagcagcagcctcaccacagag--gagagctcaag 2524
Db 359 TGAGACAGCAAGGAGAGAGTACTCCAGAGGCTCTACAGAGAGAGAGAGAGAGAG 418
Qy 2525 ttccatagcttccaaagtgtgagatcagctatgttagagcagcagctgaggaacc 2584
Db 419 CCCACATGCTTCCAAAGTGAATTCCTTACATTGAGAGCTCCACGAGAGAAACC 478
Qy 2585 tcaagactc-gtgaagagacattatcgactttgacacacca 2631
Db 479 TTGAGACTCTGTTGGAAAGGCCCTTATCTTACTTAAACACCACTTA 526

```

Search completed: February 23, 2002, 23:36:01
 Job time: 9362 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:36:01 ; Search time 6550.43 Seconds
(without alignments)
13466.612 Million cell updates/sec

Title: US-09-171-553B-3

Perfect score: 8209

Sequence: 1 gtggtgacgactgtggcc.....aaaaaaaaaaaaaaaaaaaa 8209

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_htc:*

10: gb_est1:*

11: gb_est2:*

12: gb_htc:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600	7.3	2085	12 AK014027	AK014027 Mus muscu
2	594.6	7.2	857	11 B1184013	B1184013 UNL-P-FN
3	552.4	6.7	2096	12 AK011989	AK011989 Mus muscu
4	473.6	5.8	549	10 AW657531	AW657531 110922 MA
5	472	5.7	2935	12 AK004927	AK004927 Mus muscu
6	420.6	5.1	467	11 BF441465	BF441465 257856 MA
7	419.8	5.1	466	11 BF441468	BF441468 257862 MA
8	419.4	5.1	465	11 B1399234	B1399234 MI-P-AV1-
9	404.2	4.9	801	13 BH063533	BH063533 RPCI-24-3
10	401.6	4.9	452	11 BF441466	BF441466 257857 MA
11	385	4.7	576	13 AZ730616	AZ730616 RPCI-24-1
12	383.6	4.7	564	11 B1182930	B1182930 UNL-P-FN-

ALIGNMENTS

RESULT 1

AK014027
LOCUS AK014027 2085 bp mRNA 05-JUL-2001
DEFINITION Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110009E22, full insert sequence.
ACCESSION AK014027 GI:12851651
VERSION AK014027.1
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 13 days embryo head cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2085)
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2085)
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (bases 1 to 2085)
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

13	380.8	4.6	864	13 BH056750	BH056750 RPCI-24-9
14	380.4	4.6	729	13 AZ318021	AZ318021 IM0036004
15	373.2	4.5	673	13 AZ329067	AZ329067 IM0053G10
16	371	4.5	447	11 B1401063	B1401063 MI-P-CP0-
17	367	4.5	679	13 BH121259	BH121259 RPCI-24-3
18	366.6	4.5	646	13 AZ967275	AZ967275 2M0238G10
19	358.6	4.4	602	11 BF712162	BF712162 MI-P-E6-a
20	355.4	4.3	564	11 B1181099	B1181099 UNL-P-FN-
21	355.4	4.3	685	13 AZ636636	AZ636636 IM0495002
22	354.8	4.3	564	11 B1186129	B1186129 UNL-P-FN-
23	352.4	4.3	644	13 BH124013	BH124013 RPCI-24-2
24	352	4.3	376	11 BF441469	BF441469 257863 MA
25	351.4	4.3	690	13 AZ627566	AZ627566 IM0469L20
26	349	4.3	799	11 B1107577	B1107577 602894023
27	348.8	4.2	705	13 AZ337037	AZ337037 IM0067L03
28	343.8	4.2	598	13 AZ959941	AZ959941 2M0227K18
29	343.2	4.2	791	13 BH090616	BH090616 RPCI-24-3
30	341.6	4.2	842	13 BH106844	BH106844 RPCI-24-3
31	340.2	4.1	789	13 AZ715956	AZ715956 RPCI-24-1
32	340.2	4.1	975	11 BF784826	BF784826 602111174
33	338.8	4.1	755	13 AZ715272	AZ715272 RPCI-24-1
34	338.6	4.1	601	13 AZ650638	AZ650638 IM0520L20
35	338	4.1	617	13 AZ420279	AZ420279 IM0198K05
36	336.8	4.1	653	13 BH083773	BH083773 RPCI-24-2
37	336.6	4.1	632	10 AL134888	AL134888 DKFZP762K
38	336.6	4.1	657	13 BH096361	BH096361 RPCI-24-3
39	336.4	4.1	813	11 B1152889	B1152889 602918048
40	336.2	4.1	682	13 AZ980516	AZ980516 2M0257006
41	335	4.1	700	13 AZ827524	AZ827524 2M0104E02
42	333.2	4.1	689	13 AZ897994	AZ897994 RPCI-24-1
43	331.4	4.0	817	13 AZ717731	AZ717731 RPCI-24-1
44	330.8	4.0	692	13 AZ769626	AZ769626 IM0570J04
45	330	4.0	688	13 AZ912534	AZ912534 RPCI-24-1

Db 1036 ATTCTCATCAGCTGATCTCTATTAAGTAACTGGAA-----ATCTTACTTTTTCAGAAAAATCC 1087
 QY 1292 ccaacgcctcacgggttggtgagtccttatgtctctcacagcctacttggtgatga 1351
 Db 1088 AGCTGGACTACACAGCTGGTGGAGTCAATATGTTATCTCATCAGCCCACTTAGGATGA 1147
 QY 1352 ttgtcaacagctgtgcagacactcttcacaaaccgagagcgtgagagaaattctgttaga 1411
 Db 1148 TTGCAACAGCTTTTGATGCTTCTTTTACCACAGAGAAAAAGAAAGATCTCTTAAG 1207
 QY 1412 ggttagaaaaatgttctctgggcccagggccacacacagctgtcgaataatgagattga 1471
 Db 1208 GGCTCAGAAGACTGTCTCAGGGGCAAAACAGGGCCCCCAACCACTGTCCAACAAGATAAA 1267
 QY 1472 catgggatttcccttgactcgcgcctgttgagactcaaacacgcgtgaaggttaggagag 1531
 Db 1268 CGTGGGTTCCCTCTCGATAGACAGACTGGGACTCAACACCGCTCAAGGTAGGGAGAG 1327
 QY 1532 cttgaaaaatctatgcagggctctgtggtggtctccggggcgctcgaagacggccac 1591
 Db 1328 GTTGACAGTTTACTGCGGGCTCTAGTGCAGGTCTCAAGGGAGCGCAAGATGCCCTC 1387
 QY 1592 taatttgctaaaggttaagagagtgatgagggaccgaaacgaacctccctcgttatctt 1651
 Db 1388 CAATCTGGCAAAAGGTAAAGAGACTATGCAGGGGCCAGCTGAACCCCTCTCTTTTCT 1447
 QY 1652 tgagagctcatggaagccttcagcggttcaccccttttgatctacctcagagccca 1711
 Db 1448 AGAGAGATTAATGAGGCTTACGGAGCGCTACACTCTCTTTGACCCGACCTCGAGGGTCA 1507
 QY 1712 gaagcctcagtgccctgcttcaattgggcagtcggtctggtatcaggaagaacct 1771
 Db 1508 ACAAGCCGCGGTAGCTATGCTTTTATAGGCAGTCAATAGCAGACATTAAGCAAGTT 1567
 QY 1772 tcaagactgaaggttacagagagctgagttacgtatctagtgagagagcagagaa 1831
 Db 1568 ACAGAGACTGGAAGGTTTACATGCCATGCTTACAGG----ACTTAGAGAGCGCAGAGAA 1623
 QY 1832 ggtgtattacagagggagacagagagagagagagagagagagagagagagagagag 1891
 Db 1624 AGTGTTTACAGAGGAGAGAAACAGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1681
 QY 1892 agaaaggggagaaagcgtgtatagacggcgaagagaaatttgactaatatcttgccgc 1951
 Db 1682 AAGAAAGAAAAATAGGAGGATCAGAGCGCAGGAGAGAGAAATTTAAGCAGGATCCTGGCTGC 1741
 QY 1952 agtgggtgaaggaagacagcagcagcagcagcagcagcagcagcagcagcagcagc 2011
 Db 1742 AGTAGTAGAGAGCGGTGAGGGAGAGGTAGATCTGAGAGTAGACAAACAGGCTACCTGGG 1801
 QY 2012 c-----cctagacagtcaggggaacctgggcaatagaccctcactgcacagagacca 2062
 Db 1802 CAACAGGGAACCTAGACCTTGTGGAGAGAAACCCGTTGAGAAACTCTGGAAGAGGACCA 1861
 QY 2063 gtgctgctatttgaagaaaaagacactgggcaaggaactcccccaagaggggaaacaa 2122
 Db 1862 ATATGCTATAGTCAAGAGAGAGGACATTTGGCGCGGAGAAATCCCATGAAGAGTGGGAAA 1921
 QY 2123 agaacccaaggtccctagctctagaagaagataaagattaggggagacggggttcgagacc 2182
 Db 1922 TGCCCTAAGGTATGATCTCTGAAGATGAT---GACTAGGGGAAATGGGGCTCAGACT 1978
 QY 2183 cctcccccagcccaggttaactttgaaggt-ggaggggcaacacagttgattccctggttg 2241
 Db 1979 CCTCCCAAGCCTAGGTTAACCCTGACTGTGGGAGGGGACTCCTATGACATCTTAGTAT 2038
 QY 2242 ataccggagcggagcattcagtgctgtacacaccattaggaataact 2287
 Db 2039 ATACCGGGGTAGAACACTCTGTGTTGAAACAACCATTTGGGGAAGCT 2084

RESULT 2
 B1184013

LOCUS B1184013 857 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-cf-b-12-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 ACCESSION UNL-P-FN-cf-b-12-0-UNL 3', mRNA sequence.
 VERSION B1184013
 KEYWORDS B1184013.1 GI:14658422
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 857)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message.
 Seq primer: M13 -29
 POLYA=No.

FEATURES
 source
 Location/Qualifiers
 1..857
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection
 Lines"
 /db_xref="taxon:9823"
 /clone="UNL-P-FN-cf-b-12-0-UNL"
 /clone_lib="ADULT"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
 library is a normalized library representing porcine
 ovarian follicles, ranging between 2.0 to 10.0 mm in
 diameter, collected during 7 days of the follicular phase
 of the pig estrous cycle. This library was derived from
 the library UNL-P-F2. The tag is a string of 5-6
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996.
 TAG_SEQ=None found"
 BASE COUNT 152 a 217 c 248 g 236 t 4 others
 ORIGIN

Query Match 7.2%; Score 594.6; DB 11; Length 857;
 Best Local Similarity 85.6%; Pred. No. 3.3e-110;
 Matches 773; Conservative 0; Mismatches 51; Indels 79; Gaps 7;

QY 2 tgggttacgactgtggcccccagcgccttggaataaaatcctctgctgtttgcatca 61
 |||||
 Db 12 TGGCTTAGACTGTGGGCCCCAGGTGTCTTGGAAATAAAATCCCTCTTCTGTTTGCATCA 71
 |||||
 QY 62 agacgcctctcgtgagtgattgggtgctcctctccgagcccgacgaggggatt 121
 |||||
 Db 72 AGACCGCTTCGTCGTAGTGATTTGGGGTGTGCCCTCTCTCCGAGTCCCGACGAGGGGATT 131
 |||||
 QY 122 gttcttactggccttcttatttggtgcttggcggggaatccctcgacacaccccttac 181
 |||||
 Db 132 TTCCTTCTGTCGCTTTCATTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 191
 |||||
 QY 182 acccgaaacgcagcttgaggaagaagggatccctcttggaacgtgtgtgt-gtgcggc 239
 |||||
 Db 192 ACCCGAAGACCGACTTGGAGTAAAGGATCCCTTTTGGAAATGTGTGTGTGTGTGTGTGTGT 251
 |||||
 QY 240 cggcgctctctgtctgagtgctgtttcttcggtgatgcgccttcttcggtgacgtgtcc 299


```

|||||
Db 252 CGCGCTCTCTGTTCTGAGTGTGTTCTTCTGGATCGCGCTTTCGGTTTCAGCTGTCC 311
|||||
Qy 300 tctcagaccgtaagactggaggactgtatcagacagactgctagaggatcacagct 359
|||||
Db 312 TCCACAGACCGTAAGACTGGGGAGCTGTGATCAGTAGACGCTGCTAGGAGGATCACAGCT 371
|||||
Qy 360 gccacctggggagcccccggaggtgggagagccaggagcgcctggtggtctctac 419
|||||
Db 372 GCGGCCCTGGGGACGCCCGGAGGTGGGAGAGCCAGG--TGCGTGTGCTCCTCCTTC 429
|||||
Qy 420 tctcagaccgagaccaggtctgtgttgaaagcgaagattccccctcccgccgtcc 479
|||||
Db 430 TCTCGCTCAGAGACCGAGTCTGTGTGTAAGCGAAGACTTCCCGCTCCCGCCGCT 489
|||||
Qy 480 gactctttgctctgttggaaagcgcggacgggtcggtgctgtgactctgtgtt 539
|||||
Db 490 GACTCTTTGCTCTGTGTGGAAGCGCGGAGCGGTCCGCTGCTCGATCTGTGTT 549
|||||
Qy 540 tctgtctgtgtgttctgttctgtctgtctgtctgtctgtctgtctgtctgtctgt 599
|||||
Db 550 TCTGTGCTGTGTCTTGTCTTATGCTGCTTCTCAACCAATTTGGACTGAAGTTAAATCCAG 608
|||||
Qy 600 gtgacta-----cccccttagttgactctgtctgtctgtctgtctgtctgtctgt 652
|||||
Db 609 GTGAGCACCCCGCCCGCTTATGCTGCTTATGCTGCTTCTCAACCAATTTGGACTGAAGTTAAATCCAG 668
|||||
Qy 653 gctcataattgtcagttcagtttaagaaagggaccttggcagacttctgtcctctga 712
|||||
Db 669 GCTCATATTTGTTCAGTTCAGTTAAGAA----- 698
|||||
Qy 713 atggccaacattcgatgttggatggccatcagaggggaccccttaattctgaaattatcct 772
|||||
Db 699 -----GGGACCTTAAATCTGAGATTATCCT 724
|||||
Qy 773 gctgttaaggaactaattttcagactggaccggctctcatctgtatgagggaccccta 832
|||||
Db 725 GGCTGTNAAGCAATATTTTCAGACTGGACCGCGCTCATCTGCTGATCAGGAGCCCTA 784
|||||
Qy 833 tatctctagtgagcaagatttggcagaagactctccgcctgggttaaacatcgctctaa 892
|||||
Db 785 TATCTTAGATGGCAGATGG---CAGATGCTCTCNGCATGGATTAAACCATGGCTAAT 841
|||||
Qy 893 taa 895
||
Db 842 AAA 844

RESULT 3
AK011989
LOCUS
DEFINITION Mus musculus 2096 bp mRNA HTC 05-JUL-2001
library, clone:2610305J24, full insert sequence.
ACCESSION AK011989
VERSION AK011989.1 GI:12848454
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cdna to mRNA,
clone.lib:RIKEN full-length enriched mouse cdna library
clone:2610305J24.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2096)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE High-efficiency full-length cdna cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (bases 1 to 2096)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cdna libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
20499374
MEDLINE 11042159
REFERENCE 3 (bases 1 to 2096)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530913
MEDLINE 11076861
REFERENCE 4 (bases 1 to 2096)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cdna collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2096)
REFERENCE 5 (bases 1 to 2096)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
FEATURES
source
Location/Qualifiers
1. 2096
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:1905349"
/db_xref="MGD:MGI:1919770"
/clone="2610305J24"
/clone_lib="RIKEN full-length enriched mouse cdna library"
/dev_stage="10 days embryo"
688. >2096
/note="putative"
/codon_start=1
/protein_id="BAB27961.1"
/db_xref="GI:12848455"
/translation="MGQTVSTPLSLTKDHWTDVARGQNLVSKVKKKPMWTCSSWPLP
VFGGWPAGTEFYLTIRAVKAIQFQEGSHDQOQYIMVMDLARYPPPPVPPVPLP
PLRPGTKILAIQENGKPKPLGRDDDRSTPTVKPKPIYSEIEEPWPDPPTQPPP
YAPQPSAPSGPLPOAPAGGPGSTGTSRGVTPGPDSTVALPLAIGAPPADP
NSLQPLQYWPFSDDLNNKAHPFSENPACLTGLVESLMYSHOPTWDDCOOLQTL
FTTEERERILLEARKNVRGAPVOTPAEIDEGFELTPRPWDYNTASGRRLSNYRR
VLVGLRGAARQPTNLAKRVNMQGATEPVSFLERMEAYRIYTPDPTSQGRASR
IMAFIGQSAPDIRKQLRIEGLQDYYTIRDVREAKEYVHRETTEDKKLEREKREE
EDRRRRQKVKLTRLAAGV"
BASE COUNT 500 a 527 c 577 g 492 t
ORIGIN
Query Match 6.7%; Score 552.4; DB 12; Length 2096;
Best Local Similarity 63.5%; Pred. No. 1.1e-101;

```



```

/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 145 a 144 c 121 g 139 t
ORIGIN

Query Match 5.88; Score 473.6; DB 10; Length 549;
Best Local Similarity 92.78; Pred. No. 1.1e-85;
Matches 508; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

Qy 6483 ctccccaggttctctgttaagacagacagacagactctcagctctcatccaggagcttt 6542
Db 2 CTCCCCCGGTGTTCTGTGTAGACAGGACAGAGACTCTTCAGTCTCAATCCAGGAGGTTT 61

Qy 6543 ccaagccatcaactcaccgacctgatgcacactctcttctgttggtttgtctatctc 6602
Db 62 CCAAGCCATCAACTCCACCGACCCCGATGCCAGTTCCTGTGTGGCTTGTCTATCCTC 121

Qy 6603 agggcctcttattatgaggggagtgcgttaagaagaaaaatcaatgtgacaaagagca 6662
Db 122 AGGGCTCTTATTATGAGGAATGGCTAGAGAGGGGAAATTCATTCAGTCAATGACCAAGAGCA 181

Qy 6663 tagaaatcaatgtagtggtggtccgaaataagcttaccctcactgaagttccgggaa 6722
Db 182 TAGAAATCAATGTACATGGGGGTCCCGAAATTAAGCTCACCTCACTGAAGTTCCGGGAA 241

Qy 6723 ggggacatgcatgagaaagctccccatccccacacacaccccttgctatagctgtggt 6782
Db 242 GGGACATGTCATAGGAAAGCTCCCCATCCACCAACACCTTTGCAATAGTACTATGTT 301

Qy 6783 ttatgagcagcctcagaaatcagttatttagtacctggtttataaacaggttggtggcatg 6842
Db 302 TTATGAGCAGCCCTTAGAAATCACTATTAGTACCTGGTTATTAACATGATGGTGGGCATG 361

Qy 6843 caactactgggttaacccctgtgttccacctcagctcttccacaaatccaaagattctg 6902
Db 362 CAATACTGGGTAACTCCCTGTGTTTACACACAGTCTTTAACCAATCCAAAGATTCTG 421

Qy 6903 tgcattgtccaaatcgtccccagtggtactaccactcagtgaggaagtgtccttgatga 6962
Db 422 TGTCATGTCAGCTCGTCCCTGGTACACTACCTACCTCCTGAGGAAGTAGTCATCGATGA 481

Qy 6963 atatactatcggtataaccgacaaagaaagaaacccgtatccctt-accctagtctaa 7021
Db 482 GTATGACTATCGGCTACCCGCTCAAAAGAGAGCCCGTGAACCTTAACCCCTAGCCGTTA 541

Qy 7022 tgcctgga 7029
Db 542 TACTAGGA 549

RESULT 5
AK004927
LOCUS 2935 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300007C21, full insert sequence.
ACCESSION AK004927
VERSION AK004927.1 GI:12836484
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone:1300007C21.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2935)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning

METHODS in enzymology. 303, 19-44 (1999)
99279253
10349636
2 (bases 1 to 2935)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
20499374
11042159
3 (bases 1 to 2935)
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530913
11076861
4 (bases 1 to 2935)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2935)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
15' GAGAGAGAGCGCCCAACTCGAGTTTTTTTTTTTTTTT 3', cDNA was
prepared by using triose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence[5',
GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.
Location/Qualifiers
1. 2935
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1897042"
/db_xref="MGI:1914777"
/clone="1300007C21"
/sex="male"
/tissue_type="liver"
FEATURES
source

```



```

QY 2504 accatcactgtgtgacccctccaattagatgaatgatcgaactatatttcccccaagt 2563
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2547 GCCCTGCAAGTCTGACCCCTAAACATAGAAGATGATGCTGCTACATGAGACCTCAAC 2606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2564 aaagcctgatcaagatatcagt- --cctgggttgagcaggtttccccaaagcctgggcaga 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2607 AGAGCCGGATGTTCTCTAGGGTCCACCTGGCTTCTGATTTCGCCAGGCCCTGGCGGA 2666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2621 aaccgagggatgggttggaagaagaattccccacaggttattcaactaaagccag 2680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2667 AACTGGGGGACTGGGACTGGCAGTTGCCCAAGCGCCTCTGATTATACCTCTGAAGGCAAG 2726

RESULT 6
BF441465 467 bp mRNA EST 01-DEC-2000
LOCUS 257856 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF441465
VERSION BF441465.1 GI:11501629
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 467)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
JOURNAL EST discovery in swine
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 86 row: C column: 8
Seq primer: ATTAGCTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..467
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 115 a 118 c 127 g 107 t
ORIGIN
|||||
QY 1299 ctacagggttggtgagtccttattttctccacagcctctgggatgtttcga 1358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CTACAGCGGTAGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATTCAG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1359 cagctgtcagacactcttcaaacaccgagcagcagagagaattctgttagagcctaga 1418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CAGTGTGTCAGACACTCTTCACACCGAGGAGCGAGAGAGAATTCGTGAGGCTAGA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1419 aaaaatttctctggggcagcgccaccacacagcttgcacaaatgagattgacatggga 1478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 121 AAAAATGTTCTCTGGGCGCGACGGCGACCCACCGAGTTGCAAAATGAGATTGACATGGGA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1479 ttcccttgactcctcccggttggaactacaacacgctgaagtagggagagcttgaaa 1538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 TTTCCCTTGACTCGCCCGGATGGGACTTACAACACGCGCCGAGTAGGAGAGCTTGAAA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1539 atctatcgccaggctctggtggtgggtctccggggcgccctcaagacgcccactaaattg 1598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ATCTATCGCCAGGCTCTGGTGGCGGCTCTCGAGGCGCCTCAAGACGCGCCCAATATTG 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1599 gctaagtaagagaggtgatgcaggagcccaaacacccctccctcggtatttcttgagag 1658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GCTAAGGTAGAGAAATCATGCAGGACTGAATGAACCCCTCTGTCTTTCTTGACAGG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1659 ctcatgaaccttcagcgggttcacccctttgatcctcctcagagcccgaaagcc 1718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 CTCTTGAAGCCTTCAGCGGCTACACCCCTTTTGATCCCACTCAGAGCGCTAAAAAGCC 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1719 tcatggccctggccttcattgggcagtcggtcctcgatcaggaa 1765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TCAGTGGCTTTGGCCTTTATAGACAGCTCAGCCTTGATATTAAAAA 467

RESULT 7
BF441468 466 bp mRNA EST 01-DEC-2000
LOCUS 257862 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF441468
VERSION BF441468.1 GI:11501632
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 466)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
JOURNAL EST discovery in swine
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 86 row: D column: 8
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..466
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 114 a 118 c 127 g 107 t
ORIGIN
|||||
QY Match 5.1%; Score 419.8; DB 11; Length 466;
Best Local Similarity 94.2%; Pred. No. 9.3e-75;
Matches 436; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1299 ctacacgggggtgtgagtccttattgttctctcaccagcctactctggtgattgtctcaa 1358
 Db 1 CTCACGGCGTGTAGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATTCTCAG 60
 QY 1359 cagctactcagacactcttcacacccgagagcagagagattctgttagagctaga 1418
 Db 61 CAGCTGCTGAGACACTCTTTCACACCGAGGCGGAGAGAAATCTGTTAGAGGCTAGA 120
 QY 1419 aaaaaatttctctgggcccagggcagccaccacagcttgcaaaatgagattgacatggga 1478
 Db 121 AAAAAATGTTCTTGGGCGGACGGCGACCCACGCTGTCACAAATGAGATTGACATGGGA 180
 QY 1479 ttctccttgactgcgccgggttgaggactacacacacggtgaaagtgaggagcctgaaa 1538
 Db 181 TTTCCCTTGTACTGCCCGGATGGGACTACACACGCGCGAAGGTAGGAGAGCTTGA 240
 QY 1539 atctatgccagactctgttggcggtctccggggcgccctcaagacgcccactaattg 1598
 Db 241 ATCTATGCCAGGCTCTGCTGCGGGTCTCCGAGGCGCCTCAAGACGGCCCTCAATATTG 300
 QY 1599 gctaaaggttaagagaggtgtatgcagggaccgaaacacccctccctcggtattcttggag 1658
 Db 301 GCTAAGGTAGAGAAATGATGACGAGGACTGAATGAACCCCTCTGTTTCTTGAGAGG 360
 QY 1659 ctcatggaacccctcagggggttcaacccctttgtactactcagagagccagaaagcc 1718
 Db 361 CTCTTGAAGCCTTCAGGCGGTACACCCCTTTTGTATCCACCTCAGAGGCTTAAAAAGCC 420
 QY 1719 tcaatggccctgcctcattggcgagtcaggtcggtctctggatatca 1761
 Db 421 TCAGTGGCTTTTGGCTTTATAGACAGCTCAGCCTTGGATATTA 463

RESULT 8

BI399234/c
 LOCUS BI399234 465 bp mRNA EST 14-AUG-2001
 DEFINITION MI-P-AY1-nrp-a-02-0-UI.s1 MI-P-AY1 Sus scrofa cDNA clone
 MI-P-AY1-nrp-a-02-0-UI 3', mRNA sequence.
 ACCESSION BI399234
 VERSION BI399234.1 GI:15178301
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 465)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704477
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildeer Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: ktuggle@iastate.edu
 Oligo-dt track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
 University of Iowa Clone distribution: clones will be available
 through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=No.

FEATURES
source

Location/Qualifiers
 1..465
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-AY1-nrp-a-02-0-UI"
 /clone_lib="MI-P-AY1"
 /lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-AY1
 library is normalized library derived from the MI-P-AY0
 library, ultimately derived from placenta tissue. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 http://pigest.genome.iastate.edu/. The procedure used to
 create this library has been previously described (Bonaldo
 , Lennon and Soares, Genome Research 6: 791-806, 1996)
 TAG_SEQ=None found"
 BASE COUNT 137 a 122 c 100 g 106 t
 ORIGIN

Query Match 5.1%; Score 419.4; DB 11; Length 465;
 Best Local Similarity 95.4%; Pred. No. 1.1e-74;
 Matches 432; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
 QY 493 gcttgtgaagacgagcggcgctgctgtctgtgactctgttggttctctctgtgtg 552
 Db 465 GCTTGTGGAAGACGTGGACGGGTCTGCTGTGATCTGTGTGTTTCTGTTTGTGTG 406
 QY 553 tcttgtctgtgctccttctctacagttttaataatggacagacagctactaccccc 612
 Db 405 TTTTGTGCTGTGTGCTCTGTCTACAGTTTAAATATGGACAGCGGTGACGCCCTC 346
 QY 613 ttagtctgactctgacacattgactgaagttagatccagggctcataattgtcagttc 672
 Db 345 TTAGTTTGAATCTCGACCATTTGGACTGAAGTTAAATCCAGGGCTCATATTTGTCAGTTC 286
 QY 673 aggttaagaagggaccttggcagacttctgtgcctctgaatggccaacattcagatgtg 732
 Db 285 AGGTTAAGAAGGACCTTGGCAGACTTCTGTGTCTCTGTAATGGCCGACATTCGATGTG 226
 QY 733 gatggccatcagaggggacctttaattctgaaattatctgctgtcttaagaacattt 792
 Db 225 GATGGCCATTCAGAGGGGACCTTTAAATCTTGATATTATCTGGCTCTTAAAGCAATATT 166
 QY 793 ttcagactggacccggctctctcactctgacagagccctctatctctacgtggcaagatt 852
 Db 165 TTCAGACTGGACCCGGCTCTCATCCGATCAGGACCCCTATATCTTACGTGGCAAGATT 106
 QY 853 tggcagaagactcctccgcctgggttaaacatggctgctgctgctgctgctgctgctgct 912
 Db 105 TGGCAGAGGATCTCCGCCATGGGTAAACCATGCTGTAATAGCCAAAGCAAGCCAGTC 46
 QY 913 cccgaatcctggctcttggagagaaaaacaaac 945
 Db 45 CCCGAATCTGCTCTTGGAGAGAAAAAAAC 13

RESULT 9

BI063533
 LOCUS BI063533 801 bp DNA GSS 18-JUL-2001
 DEFINITION RPCI-24-370H8.TJ RPCI-24 Mus musculus genomic clone RPCI-24-370H8,
 DNA sequence.
 ACCESSION BI063533
 VERSION BI063533.1 GI:14878283
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 801)
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-370H8.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research

Db 61 CAGCTGCTGCAGACACTCTTACAAACGAGGAGCGAGAGAAATCTGTTAGAGGTAG- 119
 QY 1419 aaaaatttctggggccgagggcgagccacgagtgcaaatgagattgacatggga 1478
 Db 120 AAAAAATGTTCTGGGGCCGAGCGGCGAGCCACGAGTTGCAAAATGAGATTGACATGGGA 179
 QY 1479 ttcccttgactgcgcggcttgggactacaacacgctgaagtgaggagagcttgaaa 1538
 Db 180 TTTCCCTTGACTCGCCCGGATGGGACTACAAACGCGCGAAGGTAGGAGAGCTTGAAA 239
 QY 1539 atctatgcagcagctctggtggtggtctccggggcgctcagagcggccacataattg 1598
 Db 240 ATCTATGCCAGGCTCTGTTGGGGGCTCTCCGAGGGCGCTCAAGACGGCCCACTAATTG 299
 QY 1599 gctaaggttaagagagtgatgagcaggaacgaacacccctccgttattcttgagag 1658
 Db 300 GCTAAGGTAAAGAAATGATGCGAGGAGCTGAATGAACCCCGCTCTGTTTCTTGAGAG 359
 QY 1659 ctcatggaagccttcagcgggcttcaaccccttttgatcctaccctcagagccagaaagcc 1718
 Db 360 CTCTTGAACCTTCAGCGGCTACACCCCTTTTGATCCACCTCAGAGGCGCTAAAAAGCC 419
 QY 1719 tcaagtgccctggccttcattggcagtcggc 1750
 Db 420 TCAGTGGCTTTGGGCTTTATAGGACAGTCAGC 451

RESULT 11

AZ730616 576 bp DNA GSS 25-JAN-2001
 LOCUS RPCI-24-149D1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-149D1,
 DEFINITION DNA sequence.
 AZ730616
 ACCESSION AZ730616.1 GI:12491859
 VERSION
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 576)
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-24-149D1.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 149 row: D column: 1
 Seq primer: SP6
 Class: BAC ends.

FEATURES
source

Location/Qualifiers
 1. 576
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-149D1"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the
 BamH1 sites using MboI partially digested male C57BL/6J
 DNA.^a

BASE COUNT 161 a 139 c 121 g 155 t
 ORIGIN

Query Match 4.7%; Score 385; DB 13; Length 576;
 Best Local Similarity 82.5%; Pred. No. 1e-67;
 Matches 453; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

QY 5040 ggaataacctaaagtgataggggtcagacaattgtccagcttcttggtgcccagtgaaagtcag 5099
 Db 1 GCATACCCCAAGGTAAATCGGGTCCATTAATGAGCCTGCTTTGTTGGCCAAAGTAAGTCAG 60
 QY 5100 ggactggccaaagatattggggattgattggaacctgcattgtgcatacacagaccccaagc 5159
 Db 61 GGACTAGCCCAAGATATTGGGGATTGATTAGAAGTTACATTGTGTATACAGGCTCCAAAC 120
 QY 5160 tcaggacaggttagagagatgaatagaaccattaaagagacccttaacaaattgaccaca 5219
 Db 121 TCAGGACAGGTAGAGAGGATGAATAAACCATAAAGAGACCCCTTACCAAAATTG-CTGCT 179
 QY 5220 gagactggcattaatgattggtggtggtcctcctgccttctgtcttctttaggtgaggaac 5279
 Db 180 ATGACTGCCACTAATGATTGGATAGCTCTCTACCCCTTTGCTCTTCAGGGTTAGAAAC 239
 QY 5280 accctggacagttgggtgctgacccctatgaattgctctacggggggagggcccccgttg 5339
 Db 240 ACCCTGGACAGTTAGACTGAACCCCTATGAATCACTCTATGGGGGTCTCTCCATTTG 299
 QY 5340 gcagaaattgctttgcacatagtgctgctgctgttccagcctttgttctctagg 5399
 Db 300 GTAAAAATAGGCTCTGTACATAGTCTGCTGCTGCTTTCCAGCCCTTTCTCTAGT 359
 QY 5400 ctcaaggcgtcagtggtggtgagcagcagcgtggaagcagctccggagggcctactca 5459
 Db 360 CTCAGAGTGCATGGATGGTGAGACAGCAAGTATAGAGAGCAGCTCCAGAAATGCCACTCA 419
 QY 5460 ggaggagacttgcagttccacatcagcttcccaagttggagattcagttatgttagagc 5519
 Db 420 AGAGAAGACCTGGAAAGCCCCACATCTCTTCCAAAGTTATAGATTTCAGTCTACATTAGAGC 479
 QY 5520 caccgtgcagaaacctgcagacacgcgtggaaggaccttatctcgtactttgaccaca 5579
 Db 480 CTCACACAGAAACCTTTAAGACTCAGCGGAAAGCCCTTATCTGTAATCTTTACCACC 539
 QY 5580 ccaacgct 5588
 Db 540 CTAATTTCT 548

RESULT 12

BI182930/c
 LOCUS BI182930 564 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-br-h-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-br-h-08-0-UNL 3', mRNA sequence.
 ACCESSION BI182930
 VERSION BI182930.1 GI:14657339
 KEYWORDS EST.
 SOURCE Sus scrofa pig.
 ORGANISM Sus scrofa

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 564)
 AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
 TITLE Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA

Tel: 402 472 6416

Fax: 402 472 6362

Email: dpoem@unl.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized porcine ovarian follicles library

Seq primer: M13 -29

POLYA=yes.

FEATURES

source Location/Qualifiers
1. 564
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection Lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-br-h-08-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_LIB=UNL-P-FN
TAG_TISSUE=porcine ovarian follicles
TAG_SEQ=CACACT"
BASE COUNT 141 a 129 c 143 g 151 t
ORIGIN

Query Match 4.78; Score 383.6; DB 11; Length 564;
Best Local Similarity 84.68; Pred. No. 2e-67;
Matches 477; Conservative 0; Mismatches 9; Indels 78; Gaps 1;
QY 7720 cctgctgctgcagtaaataggtagaaggtcacacttcctattgttccaggcgtgcta 7779
|||||
DB 564 cctgctgctgcagtaaataggtagaaggtcacacttcctattgttccaggcgtgcta 505
QY 7780 tcttgctgaagataacaggaaatgagttgactaatacgttattctgattctgtaa 7839
|||||
DB 504 tcttgctgaagataacaggaaatgagttgactaatacgttattctgattctgtaa 445
QY 7840 aactgactggcaccatagaagaattgattacattgacagccctagtgacctatctcaa 7899
|||||
DB 444 aactgactggcaccatagaagaattgattacattgacagccctagtgacctatctcaa 385
QY 7900 ctgcaatctgctactctgccaggagccacgcagatgctggacctcggagctatttaa 7959
|||||
DB 384 ctgcaatctgctactctgccaggagccacgcagatgctggacctcggagctatttaa 325
QY 7960 aatgattggtccacagagcgggctctcgattatttaaatgattggtccatgagcgc 8019
|||||
DB 324 aatgattggtccacagagcgggctctcgattatttaaatgattggtccatgagcgc 265
QY 8020 g----- 8020
DB 264 GGGCTCCGATATTTAAATGATTTGGTCCAGGAGCGCGGCTCTCGATATTTTAAAT 205
QY 8021 -----ggctctcgatattttaaatgattggtttgtaacacagg 8061
|||||
DB 204 GATTGGTCCACGAAGCGCGGCTCTCGATATTTTAAATGATTTGGTTCGACGACAGG 145
QY 8062 ctttatttgaaacccataaaagctgtccgattccgcactcgggcccagctcctctac 8121
|||||

Db 144 CTTTGTGTGAACCCCAATAAAAGCTGTCCGACTCGGCACACGCCGCCGACTCTCTAC 85
QY 8122 cctcgctggtgtacagctgtggcccccagcgcgtttggaataaaatcctctgtgtt 8181
|||||
Db 84 CCCTGGTGGCTACGACTGTGGCCGCCAGCGCTTGGATAAAAAATCCTCTGTGT 25
QY 8182 tgcatacaaaaaaataaaaaa 8205
|||||
Db 24 TGCATCAAAAAAATAAAAAA 1
RESULT 13
BH056750 864 bp DNA GSS 18-JUL-2001
LOCUS RPCI-24-94C16.TVB RPCI-24 Mus musculus genomic clone RPCI-24-94C16,
DEFINITION DNA sequence.
ACCESSION BH056750
VERSION BH056750.1 GI:14864050
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 864)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Kroi,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
MUSE BAC End Sequences from Library RPCI-24
Unpublished (1999)
TITLE Other_GSSs: RPCI-24-94C16.TJ RPCI-24-94C16.TV
JOURNAL Contact: Shaying Zhao
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 94 row: C column: 16
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
1. 864
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-94C16"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 237 a 221 c 209 g 197 t
ORIGIN
Query Match 4.68; Score 380.8; DB 13; Length 864;
Best Local Similarity 66.5%; Pred. No. 7e-67;
Matches 577; Conservative 0; Mismatches 287; Indels 4; Gaps 2;
QY 3414 gtgaagagtttttggagcagctgatttgcagactggtccgggtttgcgacc 3473
|||||
Db 1 GTAGTGAGTTCTGGGAACGGCAGGCTTTGTAGACTCTGAATACACGGTCCCAATG 60
QY 3474 ttacagccccactctaccgcgttaacaaagggggattctctctgggctcctgag 3533
|||||

Db 365 GCGAAGGTAGGAGGAGCGCTTGTCTATCGCCGAGCTTATGCGAGGTCTTCAGAGA 424
 Qy 1575 gctcaagagagccactaatttgctaaagtaagagagtgatgacagagagcagaa 1634
 Db 425 GCCGTAGACGGCCACCAATTTGGCTTAGTAAGAGAGGTCTTGCAGGGGAGACTGAA 484
 Qy 1635 cctccctcgatcttcttctgaagagctcatggaagccttcagagcggttccacctttgat 1694
 Db 485 CCACCCCTCAGTCTTCTTGTAGCGCTCATGAGGAGCATATTAAGAGGTACACCCCTTTGAC 544
 Qy 1695 cctccctcagagagccagagagcctcagtgcccttcctcatcttgagcagtgccctg 1754
 Db 545 CCCCTCAGAGGGGAGGAGGAGCGCTTGTAGCGCTTGTAGCGCTTGTAGCGCTTGTAGCG 604
 Qy 1755 gatacagagagagagcctcagagagctggaaggttcaagagagcgtgagtgatccta 1814
 Db 605 GACATTAAGAAAAGCTTCAAGAGGTGAGGAGGCTCCAGATCATATGAGCTTCAAGATTTA 664
 Qy 1815 gtgagagagagagagagtgatgatacagagagagagagagagagagagagagagag 1874
 Db 665 GTAAAGAGAGAGAGAGAGAGTCTATCATATAGAGAGAGAGAGAGAGAGAGAGAGAGA 724
 Qy 1875 aa 1876
 Db 725 GA 726

RESULT 15

A2329067/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

FEATURES

SOURCE

1.673
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U00C1M0053G10"
 /clone_lib="Mouse 10kb plasmid U00C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 4.5%; Score 373.2; DB 13; Length 673;
 Best Local Similarity 78.1%; Pred. No. 2.5e-65;
 Matches 520; Conservative 0; Mismatches 123; Indels 23; Gaps 5;

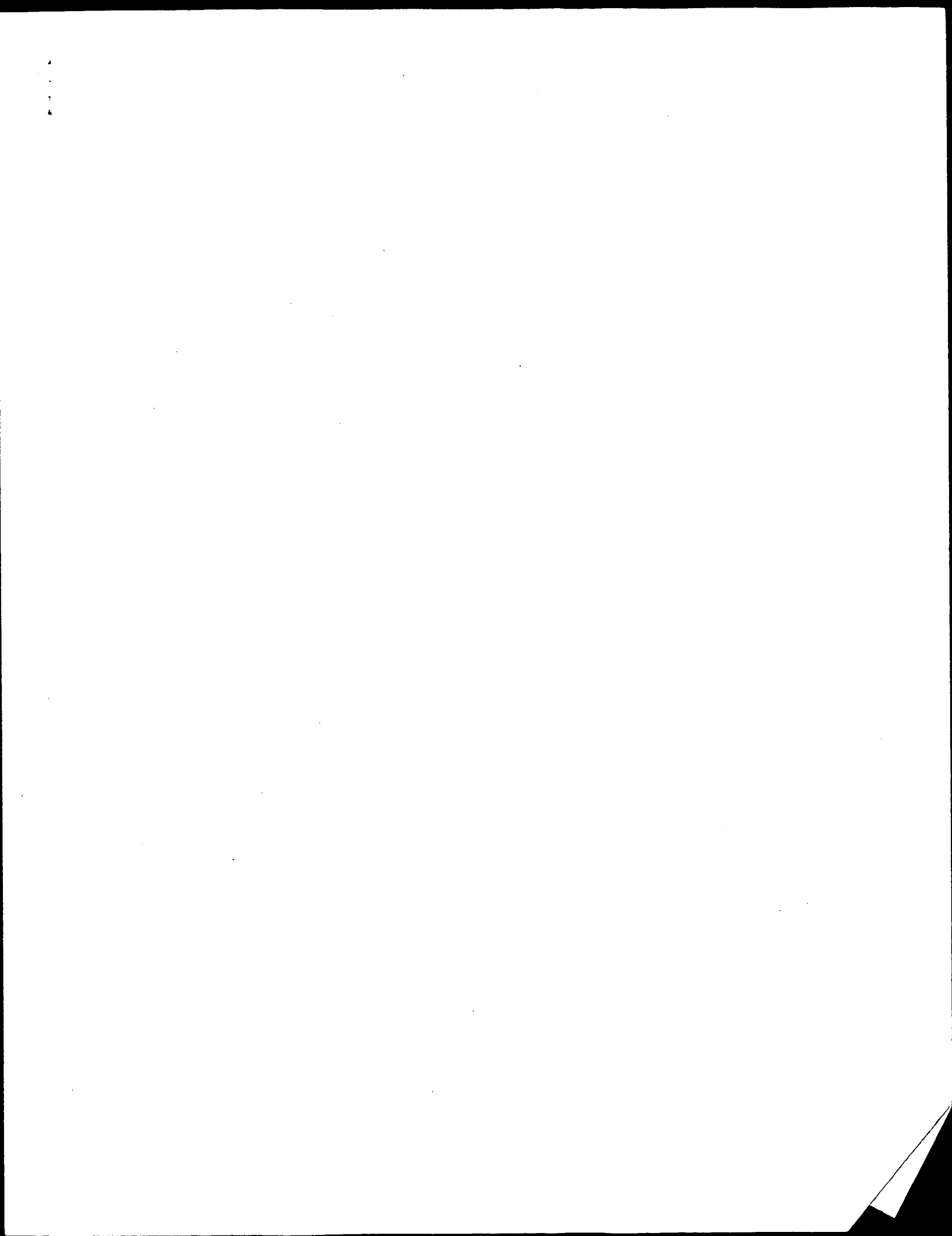
Qy 4851 ctaagggagagccaccagagcgtcctcagtgagagtgacttcagagtgagcagcgt 4910
 Db 672 CTCAGGAGTGGCTGCGCTTGAAGCTTCTGAGGAGTGAAGTTCAGAACTAGCAGCT 613
 Qy 4911 aatacagagagacacatcattatggttttgaagacaccttcagagtgagtgagcgt 4970
 Db 612 AGATACGAGCAACAGTACTTATTAGTGTATTAGATACCTTTTCAAGTGAGTACAGCT 553
 Qy 4971 taccctactaagaagaagacgttaacagctggtgagtaagaataactagagaaatttt 5030
 Db 552 TTTCCTACAGAGAGAGTGAAGC-----GGCCAAATATATCTTACACCAATCTTC 505
 Qy 5031 ccaagattggagaaacacagagagagagagagagagagagagagagagagagagag 5090
 Db 504 CCGAGGTTTGTGATACCAAGGATACAGTCAATTAAGAGAGAGAGAGAGAGAGAGAGAG 445
 Qy 5091 gtaagtcag 5150
 Db 444 GTAGTCAGGAG 385
 Qy 5151 ccccaagagcag 5210
 Db 384 CCCCAGAGCTCAG 326
 Qy 5211 ttgaccag 5270
 Db 325 TTGGCAGCTGAG 266
 Qy 5271 gtaggaacacacccctgag 5326
 Db 265 GTTAGAAGACATCCCTGAG 206
 Qy 5327 acccccccgttgcag 5386
 Db 205 GGGGCTCCACTGTAAATAATACCTCTGACATATGCTGACAGCTCTTCCAGGC 146
 Qy 5387 ttgttctcagagcag 5442
 Db 145 TTGCTTCTTAGGCTCAG 86
 Qy 5443 tccggagagcctactcagagag--gagacttgcagagagagagagagagagagagag 5500
 Db 85 TCAGAGAGGCTACTCAG 26
 Qy 5501 ttcaagt 5506
 Db 25 TTGGGT 20

Mon Feb 25 07:44:16 2002

us-09-171-553b-3.feb22std.rst

Page 15

Search completed: February 23, 2002, 23:37:03
Job time: 9424 sec



OM of: US-09-171-553b-4 to: EST:* out_format: pfs
 Date: Feb 24, 2002 8:31 AM
 About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=firmer-p2n.model -DEV=xlh
 -O/cqnt2.1/USPTO.spool/US09171553/runat-22022002-122825-11337/app-query.fasta.1.3312
 -DB=EST -OFMT=fastap -SUFFIX=feb22std.rst -GAPOP=12.000
 -GAEXP=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAEXP=7.000 -TGAP=10.000 -TGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62
 -TRANS=human4.0.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USPR=US09171553.ecgnt.1.9008 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-171-553b-4
 Query length: 524
 Database: EST.*
 Database sequences: 11351937
 Database length: 1077921985
 Search time (sec): 3515.700000

Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_hic:AK011989	+ 1553.00	1582.89	5.3e-79	2096	AK011989 Mus musculus 10 days
gb_hic:AK014027	+ 1523.00	1545.85	6.1e-77	2085	AK014027 Mus musculus 13 days
gb_hic:AK004927	+ 1223.50	1238.89	7.1e-60	2935	AK004927 Mus musculus adult ma
gb_gss:BM065533	+ 991.00	1012.63	3.1e-47	801	BM065533 RPICT-24-370H8.TJ RPICT
gb_gss:AK1318021	+ 946.00	967.63	9.8e-45	729	AK1318021 IM0036004R Mouse 10kb
gb_gss:BM121259	+ 916.00	937.70	4.6e-43	679	BM121259 RPICT-24-374G11.TV RPICT
gb_gss:AK2967275	+ 861.00	882.19	5.6e-40	646	AK2967275 2M0238G10F Mouse 10kb
gb_gss:BM124013	+ 846.00	866.80	4.0e-39	644	BM124013 RPICT-24-230P16.TV RPICT
gb_gss:AK2420279	+ 845.50	866.80	4.1e-39	617	AK2420279 IM0198K05F Mouse 10kb
gb_gss:AK2968580	+ 828.50	848.81	4.1e-38	677	AK2968580 2M0268P18R Mouse 10kb
gb_gss:AK2861103	+ 827.00	847.28	5.0e-38	677	AK2861103 2M0169D03R Mouse 10kb
gb_est2:BG089942	+ 810.50	830.89	4.1e-37	645	BG089942 mab8806.0x1 NCI CGAP
gb_gss:AK283560	+ 809.00	829.95	4.6e-37	598	AK283560 60110760F1 NCI CGAP
gb_est1:BE283558	+ 800.00	829.95	4.6e-37	601	BE283558 IM0520L20R Mouse 10kb
gb_gss:AK2708792	+ 797.50	815.83	2.4e-36	585	AK2708792 RPICT-23-238P4.TJ RPICT
gb_est1:BE283558	+ 796.00	816.91	2.4e-36	585	BE283558 601102758F1 NCI CGAP
gb_est2:BP143092	+ 793.00	810.61	9.0e-36	892	BP143092 2M0177M08R Mouse 10kb
gb_gss:AK277287	+ 787.50	806.72	1.3e-35	716	AK277287 2M011P01R Mouse 10kb
gb_gss:AK2866922	+ 783.00	803.62	1.3e-35	591	AK2866922 2M0057L01F Mouse 10kb
gb_gss:AK2799523	+ 782.00	802.63	1.5e-35	589	AK2799523 2M0057L01F Mouse 10kb
gb_gss:AK2348675	+ 782.00	802.58	1.5e-35	593	AK2348675 IM0085P17F Mouse 10kb
gb_gss:AK2970796	+ 776.50	795.95	3.6e-35	614	AK2970796 2M0244L04F Mouse 10kb
gb_est2:BF450334	+ 776.00	796.21	3.5e-35	619	BF450334 2M0277C10.X1 NCI CGAP
gb_gss:BM112876	+ 773.00	793.98	4.6e-35	552	BM112876 RPICT-24-368D12.TJ RPICT
gb_gss:BM086028	+ 764.50	784.19	1.6e-34	642	BM086028 RPICT-24-338016.TV RPICT
gb_est2:BF441468	+ 762.00	784.11	1.6e-34	466	BF441468 257865 MARC 2P1G Sus s
gb_est2:BF441465	+ 762.00	784.11	1.6e-34	467	BF441465 257865 MARC 2P1G Sus s
gb_est1:AW763885	+ 745.00	765.05	1.9e-33	588	AW763885 ut66f07.x1 NCI CGAP
gb_est1:AW763885	+ 744.00	764.37	2.1e-33	563	AW763885 2M0165A03R Mouse 10kb
gb_est1:AW763885	+ 744.00	764.37	2.1e-33	602	AW763885 2M0165A03R Mouse 10kb
gb_est2:BF450406	+ 728.00	747.59	1.8e-32	546	BF450406 ut78B07.x1 NCI CGAP
gb_est2:BF450406	+ 727.00	747.33	1.8e-32	546	BF450406 2M0165A03R Mouse 10kb
gb_gss:BM061935	+ 727.00	746.39	2.1e-32	617	BM061935 RPICT-24-370D19.TJ RPICT
gb_gss:AK2385280	+ 727.00	746.21	2.1e-32	631	AK2385280 IM0143L11R Mouse 10kb
gb_gss:AK2325348	+ 727.00	746.21	2.1e-32	631	AK2325348 IM0047N11R Mouse 10kb
gb_gss:AK2325348	+ 723.50	742.17	3.6e-32	672	AK2325348 2M0257D10F Mouse 10kb
gb_gss:AK281497	+ 718.50	736.74	7.1e-32	704	AK281497 2M0244H01.X1 NCI CGAP
gb_est1:AW214497	+ 718.00	734.61	9.8e-32	452	AW214497 257857 MARC 2P1G Sus s
gb_est1:BF441466	+ 715.50	733.28	9.8e-32	652	BF441466 uz88F01.x1 NCI CGAP
gb_est1:BF146613	+ 714.00	733.41	1.1e-31	599	BF146613 uy32B01.x1 NCI CGAP
gb_est1:BF921696	+ 713.00	730.97	1.7e-31	810	BF921696 602825515F1 NCI CGAP
gb_gss:AK2657594	+ 702.00	720.07	5.4e-31	618	AK2657594 IM0533F21R Mouse 10kb

gb_gss:AK2959168 - 700.00 719.30 6.7e-31 590 AK2959168 2M0226C14R Mouse
 gb_gss:AK2937454 - 697.50 716.23 9.9e-31 632 AK2937454 2M0195G09R Mouse
 gb_est2:BI081838 + 695.50 711.79 1.8e-30 864 BI081838 602879750F1 NCI
 gb_gss:AK2821842 - 630.50 709.15 2.5e-30 629 AK2821842 2M0094H15R Mouse
 gb_hic:AK007448 + 630.50 707.13 3.2e-30 818 AK007448 Mus musculus 10 d

seq_name: gb_hic:AK011989

seq_documentation block: 2096 bp mRNA HTG 05-JUL-2001
 LOCUS AK011989
 DEFINITION Mus musculus 10 days embryo cDNA, RIKEN full-length enriched library, clone:2610305J24, full insert sequence.

ACCESSION AK011989
 VERSION AK011989.1 GI:12848454
 KEYWORDS CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library
 clone:2610305J24.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 2096)
 Carinici, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Methods in enzymology. 303, 19-44 (1999)
 99279253
 PUBMED 10349636

2 (bases 1 to 2096)
 Carinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome research. 10 (10), 1617-1630 (2000)
 20499374

TITLE

JOURNAL MEDLINE
 11042159
 3 (bases 1 to 2096)
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carinici, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kashiwagi, K.,
 Fujimoto, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Tanaka, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multipipette sequencer
 Genome research. 10 (11), 1757-1771 (2000)
 20530913

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)

TITLE

JOURNAL MEDLINE
 11076861
 4 (bases 1 to 2096)
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 2096)
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Stehito-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
 uri:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)

1976	ATCATGAGAGAAACAGAAATAAAAACTTGAGAGAGAGAAAAGAGAG	2025
434	ArggluGluArgGluGluArgGAspArgGlnGluLysAsnLeuTh	450
2026	AAAGAGAGAGAGATAGAGAGACAGAGAGCAAGAAAAGTTTTCAGC	2075
450	rlvsllleuAlaIaVal	456
2076	TAGATCTGTGGCCGACGTA	2094
seq_name:	gb_hlc:AK014027	
seq_documentation_block:		
LOCUS	AK014027	2085 bp mRNA hlc 05-JUL-2001
DEFINITION	Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched library, clone:3110009E22, full insert sequence.	
ACCESSION	AK014027	
VERSION	AK014027.1	GI:12851651
KEYWORDS	CAP trapper.	
SOURCE	Mus musculus (strain:C57BL/6J) 13 days embryo head cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library clone:3110009E22.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)	
MEDLINE	20495374	
PUBMED	11042159	
REFERENCE	3 (bases 1 to 2085)	
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kusunagi,T., Tashiro,H., Itoh,M., Sumi.N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,Y., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneida,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuno,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.	
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer	
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PUBMED	11076861	
REFERENCE	4 (bases 1 to 2085)	
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
JOURNAL	Nature 409, 685-690 (2001)	
REFERENCE	5 (bases 1 to 2085)	
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arikawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Haraagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Imotoh,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Yamamatsu,M. and Hayashizaki,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),	

```

COMMENT
    RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
kanagawa 230-0045, Japan (E-mail: res@gsf.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAAGCAACATCCAAAGCCTCTTTTGTTCCTTTTTTGGN 3']  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 10.0 and subtraction to  
Rot = 50.0. Second strand cDNA was prepared with the primer adapter  
of sequence [5'  
GAGAGAGAGATTCTCGAGTGAATAAATTAAATCCCCCCCCCC 3']. cDNA was cleaved  
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.  
Host: SOLR.
```

```
FEATURES
     source             Location/Qualifiers
                        1..2085
                         /organism="Mus musculus"
                         /strain="C57BL/6J"
                         /db_xref="taxon:10090"
                         /db_xref="MGD:MGI:1906029"
                         /db_xref="WGD:MGI:1920342"
                         /clone="3110009E22"
                         /tissue_type="head"
                         /clone_lib="RIKEN full-length enriched mouse cDNA library"
                         /dev_stage="j3 days embryo"
                         /note="putative"
                         /codon_start=1
                         /protein_id="PAB29122.1"
                         /db_xref="GI:12851652"
                         /translation="MGEIVTPPLSLTDHMSSETRRAHNLSVEIKKGPWFPCASBPP  
TKDGWPFGSTFDLVLEFKALVFDGPSHPDOQYITVMODLYONSPPWKRPAP  
KRKGSRALAIRPKVRKDPRKEELPIEGSAIRKITPEIEGWEPWEPPPYPFLPP  
ASPRLPOLDEPGSGAGCTGVGPAAGTSGRASNOSISADGVPTDSAFYLCAVGPAADLH  
SLDTIOGWPFPSADLYNMKSIFYFRRSSWTNPQGVINVISSAHLG"
```

```
BASE COUNT      544   a       504   c       573   g       464   t
ORIGIN
```

```
alignment_scores:
          Quality: 1523.00           Length:        555
          Ratio:   3.644              Gaps:            14
Percent Similarity: 75.315         Percent Identity: 58.378
```

```
alignment_block:
US-09-171-553B-4 x AK014027 ..
```

```
Align seg 1/1   to: AK014027 from: 1 to: 2085
```

```
1 MetGlglntThrValThrProLeuSerLeuThrLeuAspHisTrpTh 17
|||||
356 AUGGACAGACTGTGCAGACCOCCTTAAGTTTGACTTAGCATCAGTC 405
|||||
17 rglUalAtgSerArGAlAhIsAnleuSevVa]GlnValLySyGlyLP 34
|||
406 TGAATATTAGCTGAGGGGCCACAATCTTTCACTGGAGAACGAAGGAGC 455
|||
34 rotTpdlnthrhEcysAlaserglUtpPrOthrPhesArva]lglyTP 50
|||||
456 CTttggcgacatttttgcttgcactgtgccacaAcAttTaTatgtggcctg 505
|||||
51 ProserGluglYthrPhesAnserrgiuleileleualvallysalaIl 67
|||
506 CCACACGAGGGGAGCTTTTGACTGACCTGATGCTTTGAAGAACCAT 555
|||||
67 ellePheginlhrglyProc]ySenhisTroksrrngluiprotYyltel 84
|:::||||| |||
```


PUBMED 11042159
 REFERENCE 3 (bases 1 to 2935)
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E., Wataniki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome research. 10 (11), 1757-1771 (2000)
 JOURNAL MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4 (bases 1 to 2935)
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL NATURE 409, 685-690 (2001)
 REFERENCE 5 (bases 1 to 2935)
 AUTHORS Atakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
 Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F.,
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
 Kojima,Y., Konno,H., Konda,M., Koya,S., Kurihara,C., Matsuyama,T.,
 Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
 Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
 Sasagaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
 Tanaka,T., Tejima,Y., Toyota,M., Yamamura,T., Yasunishi,A.,
 Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp,
 URL:http://genome.gscc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)
 COMMENT Please visit our web site (http://genome.gscc.riken.go.jp/) for
 further details.
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGAGAGAGAGCGGCCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence[5'
 GAGAGAGAGAGATCCAGACATCAATTAATTATTAACCCCCCCC 3']. cDNA was
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
 XhoI. Host: SOLR.
 FEATURES
 source Location/Qualifiers
 1..2935
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:1897042"
 /db_xref="MGI:1914777"
 /clone="1300007C21"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 BASE COUNT 669 a 860 c 773 g 633 t
 ORIGIN
 alignment_scores: Length: 563
 Quality: 1223.50 Gaps: 17
 Ratio: 3.245 Percent Identity: 48.135
 Percent Similarity: 66.963

alignment_block:
 US-09-171-553B-4 x AK004927 ..
 Align seg 1/1 to: AK004927 from: 1 to: 2935
 1 MetGlylnThrValThr.ProLeuSerLeuThrLeuAspHisTrpT 17
 639 ATGGAGACAGACCGTAACACTACCCCCCGAGACTTAACCTTCACACACACG 688
 17 hrgIuValArgSerAlaHisAsnLeuSerValGlnValLysLysGly 33
 689 GAGATGTCCAGCGCATTCATCCATCATGCTGTGGATGTGCAGAGAGAG 738
 34 ProTrpGlnThrPheCysAlaSerGluTrpProThrPheAspValGlyTr 50
 739 CGCTGGGTACTTCTTCTTCGCCGAATGGCCAACTTTCATATGGGATG 788
 50 ProSerGluGlyThrPheAsnSerGluLeileuAlaValLysAla 67
 789 GCCTCAGATGATGCTTTATTTGGGTATATCTCTCAGCTCAGTCA 838
 67 leilePheGlnThrGlyProGlySerHisProAspGlnGluProTyrIle 83
 839 GAGGTGTTTCTTCTGCTGGTCCCGACGACACCGGATCGAGTCCATATATC 888
 84 LeuThrTrpGlnAspLeuAla.GluAspProPro..... 94
 889 GTCACTGGAGGACACTCTCCAAACCCCTCTTACGACAGCTCCCG 938
 94 94
 939 TCTCCCGCGCGGCTCTTGCGCAACCTCGTCCGATCTCCCTTAC 988
 95ProTrpValLysProTrpPheAsnLysProArgLysPr 107
 989 CCGCGCCCTTACCCCTCTATTAAGCC.....AAACCTCTTAAGC 1029
 107GlyProArgIleLeuAlaLeuGlyGlu 117
 1030 CCAAGTTCCTCCATGATAGCGCGGACCTCTCATTTGACCTTCTCAGAGAG 1079
 117 ysaAsnLysHisSerAlaGlnLysValGluProSerSerSerTyrLeuPro 133
 1080 ACCCGCGCGGTAGCA.....GCACAACCTTCCTCTCT.....GCC 1117
 134 ArgAspArgGlyAlaAlaAspLeuAlaGlyThrProThrCysSerProth 150
 1118 AGAGAGGACGATGAMAAAGAGCGCGCGCCAC..... 1150
 150 rProLeuSerSerThrGlyCysCysGlnGlyThrSerAlaProProGly 167
 1151TCCGAGGTTCCTCCCTT 1169
 167 labProValGlnGlyProAlaAlaGlyThrArgSerArgArgGlyAla 183
 1170 CTCACATGATG.....TCTCGACTCGGGGAAGAGGAGGACCGCT 1207
 184 ThrProGlnArgThrAspGlnIleAlaIleLeuProLeuArgThrTyrG 200
 1208 CCGCGACCGGACCTCCACTCTCCAGGCACTTCCACTCCGATG..... 1252
 200 yProPheMetProGlyGlnLeuGlnProLeuGlnTyrTrpProPheS 217
 1253GGGGAGATGGCCAG...CTTCAGTACTGCGCTTTT 1286
 217 erSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGlu 233
 1287 CTTCTCTGACTT.....AAAATATTAACCTTCCTTTCTGAA 1326
 234 AspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisG 250
 1327 GACCCAGTAAATTTGACGGCTTGATTTGATGCTCTCATCACCCACCA 1376

```

250 nProthrtrPASPcysGlnGlnLeuLeuInthrLeuPheThrng 267
1377 GCCCAGCTGGAGAGCTGTGACAGCTGTGGGAGCTTGGACCGAG 1426
267 LgluArgLgluArgLgluLeuGlnAlaArgLysAsnValProGlyAla 283
1427 AAGAAAGACGAGGGGTGCTCTAGAGGCTAGAAAGACGAGCGGCAAT 1476
284 AspGlyArgProThrGlnLeuGlnGlnGlnLeuAspMetGlyPhePro 300
1477 GATGAGCGCCGACGCTGCTGCTATGATGATGATGATGATGATGAT 1526
300 ThrArgProGlyArgProThrGlnLeuGlnGlnGlnLeuGlnGln 317
1527 TGAAGCGCCGATGGGATGATGATGATGATGATGATGATGATGAT 1576
317 ySleThrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArg 333
1577 TCCCTATGCGCAGTGTGCTTACGGGGTCTCAAAACGCGGCGAGAG 1626
334 ProThrAsnLeuAlaLysValArgLgluValMetGlnGlyProAsn 350
1627 CCCACCAATTTGGCCAAAGGTAAAGGATAAACCCAGGAGCTTAAT 1676
350 OfProSerValPheLeuGlnArgLeuMetGlnAlaPheArgPheThr 367
1677 TCCCTCAGCGCTTTTAGAGAGACTCAAGAGGATATGCGAGTACACT 1726
367 iOpheAspProThrSerGlnAlaGlnLysAlaSerValAlaLeuAla 383
1727 CTTATGACCTTGAGAGACCCAGGCAAGCAATGCTCTATGTCATTC 1776
384 lIleGlyGlnSerAlaLeuAspLleArgLysLysLeuGlnArgLeu 400
1777 ATCTGGAGCTGTCCCGGATATCGGCGAAAGTTAGAGCGGTAAAGA 1826
400 yLeuGlnGlnAlaGlnLeuArgLgluValArgLgluAlaGlnVal 417
1827 TTTAAAGACAGACCTTAGGAGACTTAGGAGGAGAGCTGAAGAATCT 1876
417 yTTrArgArgLgluThrGlnGlnGlnGlnGlnGlnGlnGlnGln 433
1877 TTAATAAGAGAAACCCCGGAAAGAGAGAGATATCGAGAGAGAA 1926
434 ArgGlnGlnArgLgluArg..... 440
1927 ACAGAGGAAAGAGAAAGACCGCTAGGCGAGATGAGACAGAGAGAA 1976
441 ...ArgAspArgArgGlnGlnLysAsnLeuThrLysLleLeuAla 456
1977 AAGAAAGACCGGAGAGACATAGAGATGAGAGAGCTTGGGCGACTG 2026
456 aValGlnGlnLysSerSerArgGlnArgGlnArgAspPheArgLys 472
2027 TAGTAGTAGTGACAG..... 2040
473 ArgSerGlyProArgGlnSerGlnLysLeuGlnGlnGlnGlnGln 489
2041 AACACAGGCT...AGACAGGGGGGAGAG...CGAAGAGAGCGCCCA 2084
489 pLysAspGlnGlnAlaTyrCysLysGlnLysGlnLysLysLysLys 506
2085 TAAAGACCAATGCGCTACTGCAAAAGAAAGAGACATGGGCTAAGG 2134
506 ySProLysLys.....GlyAsnLysGlyProLys 515
2135 GCCCAAGAAAGACGAGAGAGAGAGAGAGAGAGAG 2169
seq_name: gb_uss:BH063533
seq_documentation_block:
LOCUS BH063533 801 bp DNA GSS 18-JUL-2001

```

```

DEFINITION   RPCI-24-370H8.TJ RPCI-24 Mus musculus genomic clone RPCI-24-370H8,
              DNA sequence.
ACCESSION    BH063533
VERSION      BH063533.1 GI:14878283
KEYWORDS     GSS.
SOURCE       house mouse.
ORGANISM     Mus musculus.
REFERENCE    1 (bases 1 to 801)
AUTHORS      Zhao,S., Nieman,M., Malek,J., Shatsman,S., Akimov,B., Levins,M.,
              Tsegaye,G., Geer,K., Kroll,M., Shwartsbeyn,A., Gebregorgis,E.,
              Russell,D., de Jong,P. and Fraser,C.M.
TITLE        Mouse BAC End Sequences from Library RPCI-24
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: RPCI-24-370H8.TV
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the mouse BAC library RPCI-24. For BAC
              library availability, please contact Pieter de Jong
              (pdejong@mail.cho.org). Clones may be purchased from BACPAC
              Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
              page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
              Plate: 370 row: H column: 8
              Seq primer: SP6
              Class: BAC ends.
              Location/Qualifiers
                source          1..801
                organism="Mus musculus"
                strain="C57BL/6J"
                db_xref="taxon:10090"
                clone="RPCI-24-370H8"
                /clone.lib="RPCI-24"
                /sex="Male"
                /cell_type="Spleen/Brain"
                /note="Vector: pPRBAC1; Site_1: BamHI; Site_2: BamHI.
                The library was cloned in the pPRBAC1 cloning vector at the
                BamHI sites using MboI partially digested male C57BL/6J
                DNA."
BASE COUNT   227 a 193 c 217 g 164 t
ORIGIN
alignment_scores:
  Quality: 991.00      Length: 271
  Ratio: 4.164        Gaps: 1
  Percent Similarity: 87.823   Percent Identity: 71.587
alignment_block:
US-09-171-553B-4 x BH063533 ..
Align seg 1/1 to: BH063533 from: 1 to: 801
215 ProPheSerSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProph 231
1 CCTTTTCTCTCTCTGATTA.TATACTTGGAAAACTAACACACCTCCCTT 49
231 eSerGlnAspProGlnArgLeuThrGlyLeuValGlnSerLeuMetPheS 248
50 CTCGAGAACCCCTCTGCTTACTGGGCTCTTGAATCACTATATGTTCT 99
248 erHisGlnProThrTrpAspAspCysGlnGlnLeuGlnGlnThrLeuPhe 264
100 CCCATCAACCCACTTGGGATGATGTCAGACAGCTTTGGCAGGTTCTTTC 149
265 ThrThrGlnGlnArgGlnArgGlnLeuLeuGlnAlaArgLysAsnValPr 281

```


258 nteuLeuGlnThrLeuPheThrThrcGluGluArgGluArgIleLeuLeuG 275
 211 GCTTTGACAGGTTCTTTCCACACAGAAAGAGAAAGAAATCCGATGG 260
 275 TAAATATGlyAsnValProGlyAlaAspGlyArgProthrgInLeuGln 291
 261 AGCGGAGAAAAATGTTCTAGGAGAGACACACACCCCTCCCTCCT 310
 292 AsnGlnIleAspMetGlyPheProLeuThrArgProGlyTyrPaspTyrAs 308
 311 AACCTGTCAGAGGCTTTCCCTTGACACGCCCAATTTGGCTAGGTAAGA 360
 308 nThrAlaGlnGlyArgGlnSerLeuLysIleTyrArgGlnAlaLeuValA 325
 361 CACCGCGGAGGTAGGGAGCGCCCTGCTCTATCGCGGACTCTAGAGG 410
 325 IagGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaValArg 341
 411 CAGCTCTCAGAGAGCGGCTAGACGGCCACCAATTTGGCTAGGTAAGA 460
 342 GluValMetGlnGlyProAsnGlnProSerValPheLeuGluArgLe 358
 461 GAGCTCTTCAGGGGAGGAGCTGACACCCCTCAGTCTTCTGAGCGTCT 510
 358 uMetGluAlaPheArgArgPheThrProPheAspProthrsSerGluAlaG 375
 511 CATGGAGGCATATAGAGGTACACCCCTTTGACCCCTGCTCAGAGGGGC 560
 375 IAlaValaSerValAlaLeuAlaPheIleGlyGlnSerAlaLeuAspIle 391
 561 AAAAAAGCCCTTACCATGGCTTCATTTGTCAGTCCGCTCCGACATTT 610
 392 ArgGlyLysLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGluLeuArgAs 408
 611 AAGAAAGACCTGCAAGGCTGGAGGGGCTCCAGATCATACGCTCCAGGA 660
 408 PheValArgGluAlaGlnLysValTyrTyrArgArgGlnThrgInGluG 425
 661 TTTTACTAAAGACAGCAGAGAAAGTCTATCATAGAGGAGAAACAGAAAG 710
 425 IuLysGlnGlnArgLys 430
 711 AGAGCGAGGAGAGAG 727

seq_name: gb_gss:BH121259

seq_documentation_block:
 LOCUS BH121259 679 bp DNA GSS 19-JUL-2001
 DEFINITION RPCI-24-374G11.TV RPCI-24 Mus musculus genomic clone RPCI-24-374G11
 , DNA sequence.
 ACCESSION BH121259
 VERSION BH121259.1 GI:14964771
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 679)
 Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akhret, B., Levins, M.,
 Tseng, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregiorgis, E.,
 Russell, D., de Jong, P., and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other GSSs: RPCI-24-374G11.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genome Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/Bac_ends/mouse/bac_end_intro.html
 Plate: 374 row: G column: 11
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. 679
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-374G11"
 /clone_11b="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI;
 library was cloned in the pTRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 194 a 167 c 174 g 144 t
 ORIGIN

alignment_scores:
 Quality: 916.00 Length: 226
 Ratio: 4.425 Gaps: 1
 Percent Similarity: 91.593 Percent Identity: 76.991

alignment_block:
 US-09-171-553B-4 x BH121259 ..
 Align seg 1/1 to: BH121259 from: 1 to: 679

214 TrrProPheSerSerAlaAspLeuTyrAsnTrrLysThrAsnHisProp 230
 1 TGGCCTTTTCTCTGATTTATACAAATGGGAAACATMACCACCCTT 50
 230 rOPheSerGluAspProGlnArgLeuThrcGlyLeuValGlnSerLeuMet 246
 51 CCTTCTCAGAGAACCCCTCTGAGACTTACTGGCTCCTTGAGTACTTATG 100
 247 PheSerHisGlnProThrrTrrPaspAspCysGlnGlnLeuGlnThrrLe 263
 101 TTCTCCATCAGACCCACTTGAGATGATTTGCACAGCTTTTGACAGTTCT 150
 263 uPheThrThrcGlnGlnArgGlnArgIleLeuGlnGlnAlaArgLysAsn 280
 151 TTTTACCACAGAGAAAGAGAAAGAAATCCTGATGAGGCGAGAAATAAG 200
 280 aLProGlyAlaAspGlyArgProThrgInGlnGlnAsnGlnIleAspMet 296
 201 TTCTGGAGAGAGGAGGACACCCACTGCTCCCTCAACCTCGTGGACGAG 250
 297 GlyPheProLeuThrArgProGlyTyrPaspTyrAsnThrrAlaGlnLys 313
 251 GCTTTTCCCTTGAACACACCCCACTGGACTACACACCGGAGAGGTAG 300
 313 gGlnSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyA 330
 301 GGGAGCGCTCTGCTATGCGCGGACTCTAGAGGAGGTCTCAGAGGAG 350
 330 IAserArgArgProThrrAsnLeuAlaValArgGlnValMetGlnGly 346
 351 CCGCTAGACGGCCCAATTTGGCTAAGTAAGAGAGGTCTTGACAGGG 400
 347 ProAsnGlnProProSerValPheLeuGlnArgLeuMetGluAlaPheAr 363
 401 CAGACTGAAACACCATCAGTCTCTTGAAGCGCTATATGAGACGATATAG 450
 363 gArgPheThrrProPheAspProThrsSerGluAlaGlnLysAlaSerValA 380
 451 GAGATACACCCCTTTGACCCCTTTCAGAGAGGCGAGAGACCGCTGATG 500


```

seq_documentation_block:
LOCUS      BH124013      644 bp      DNA      GSS      19-JUL-2001
DEFINITION RPCI-24-230P16.TV RPCI-24 Mus musculus genomic clone RPCI-24-230P16
            , DNA sequence.
ACCESSION  BH124013
VERSION    BH124013.1 GI:14967525
KEYWORDS
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 644)
AUTHORS    Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-24
            Other_GSSs: RPCI-24-230P16.T1
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pje@ngemail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
            Page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 230 row: P column: 16
            Seq primer: T7
            Class: BAC ends.
FEATURES
source     Location/Qualifiers
            1..644
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="RPCI-24-230P16"
               /clone_id="RPCI-24"
               /sex="Male"
               /cell_type="Spleen/Brain"
               /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
               RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
               library was cloned in the pTARBAC1 cloning vector at the
               BamHI sites using MboI partially digested male C57BL/6J
               DNA."
BASE COUNT  176 a      162 c      169 g      137 t
ORIGIN
alignment_scores:
            Quality: 846.00      Length: 213
            Ratio: 4.383      Gaps: 0
            Percent Similarity: 90.610      Percent Identity: 75.117
alignment_block:
US-09-171-553B-4 x BH124013
Align seg 1/1 to: BH124013 from: 1 to: 644
223 AAsTTPLVSthrAsnHisProPhesSerGluAspProGlnArgLeuTh 239
|||||
3 AATTGGAAACCTAACCCCTCCCTTCACAGAACCCCTCGACTTAC 52
|||||
239 rGlyLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspSc 256
|||||
53 TGGGGCTCCCTGAGTCATTAGTCTCCATCAACCCACTGGGATGATT 102
|||||
256 ySGlnGlnLeuLeuGlnInrLeuPheThrThrGlnGlnArgGlnArgLe 272
|||||
103 GTACAGACGCTTTGCAGGTCTCTTTTACACAGAAAGAAAGAAATC 152

```

```

273 LeuLeuGlnAlaArgLyAsnValProGlnAlaAspGlyArgProThrG 289
|||||
153 CTGATGAGAGCGGATTTTTTTTGTCTGTGGAGAGAGCGGACACCACTGC 202
|||||
289 nLeuGlnAsnGlnIleAspMetGlyPheProLeuThrArgProGlyTrpA 306
|||||
203 CCTCCCTAACCTGCTTGACGAGGCTTTTCCCTTACACCGCCCAACTGG 252
|||||
306 sPTyrAsnThrAlaGlnGluValArgGluSerLeuLysIleTyrArgGlnAla 322
|||||
253 ACTACAAACACCGCGGAGAGTGGGAGCGCTCTTCTGTATCGCCGACT 302
|||||
323 LeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAla 339
|||||
303 CTAGAGGAGAGGCTCAGAGAGCGCGTACAGCGCCACCAATTTGGCTAA 352
|||||
339 sValArgGluValMetGlnGlyProAsnGluProProSerValPheLeu 356
|||||
353 GGTAAAGAGAGGCTTGGCAGGGGAGACTGAACCACTTCACCTCTTCTTG 402
|||||
356 lUArgLeuMetGluAlaPheArgArgPheThrProPheAspProThrSer 372
|||||
403 AGCGTCTATGAGGACATATGAGATACACCTTTTGACCCCTTGCTCA 452
|||||
373 GluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyInSerAla 389
|||||
453 GAGGGGCGAGAGAGCGCGTATGACCATGGCCCTTATTTGGTCAATCGCTCC 502
|||||
389 uAspIleArgLysLysLeuGlnArgLeuGlnGlyLeuGlnGlnAlaGlu 406
|||||
503 CGACATTAAGAAAGAAAGCGTCGAGAGCGTGGAGGGCTCCAGATCATACGC 552
|||||
406 euArgAspLeuValArgGluAlaGlnLysValTyrTyrArgArgGluThr 422
|||||
553 TTCCAAGATTATGTTAAAGAACACAGAAAGTCTTCTCTAAAGGAAACA 602
|||||
423 GluGlnGluLysGlnGlnArgLysGlnLysGlnArgGlu 435
|||||
603 CGAGAGAGAGAGCGAGAGAGAGAGAGAAAGAAATGTAGAG 641
|||||
seq_name: gb_gss:AZ420279
seq_documentation_block:
LOCUS      AZ420279      617 bp      DNA      GSS      03-OCT-2000
DEFINITION IM0198K05F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
            clone UNGC1M0198K05 F, DNA sequence.
ACCESSION  AZ420279
VERSION    AZ420279.1 GI:10544292
KEYWORDS
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 617)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingley,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddun@genetics.utah.edu
            Insert Length: 1000 Std Error: 0.00
            Plate: 0198 row: K column: 05
            Seq primer: CATTGTAAACGACGCGCAGT

```


alignment_scores:
 Quality: 827.00 Length: 223
 Ratio: 4.241 Gaps: 3
 Percent Similarity: 87.444 Percent Identity: 73.991

alignment_block:
 US-09-171-553B-4 x AZ862103/rev ..

Align seg 1/1 to reverse of: AZ862103 from: 1 to: 677

```

189 AspgluileuAlaIleuProleuArgThr...GlyProPromePr 204
||||| :||| |||||||||:||||| ||| |||
660 GATTCACAGAGTTCCTGACCTAGACACATGTGGAGGCGCAGCGC 611
204 OGly...GlyInLeuGlnProLeuGlnTyrTyrProPheSerAla 220
||||| :||||| ||||||||| ||||||| :|||
610 AGGACCTAATGATTCATTCCTTACATCTGCGCTTTCTCTCTG 561
220 sPLeuTyrAsnTyrPlyThrAsnHisPropheSerGluAspProGln 236
||||| :||||| ||||||||| ||||||| :|||
560 ATTATATATTTGAAATAACCAACCTCCCTCTCAGAGAACCCCTCT 511
237 ArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGlnProThr 253
||||| :||||| ||||||||| ||||||| :|||
510 GGGCTTACTGGGCTCCTGAGTCACTTATGTTCCCATCAACCCACTTG 461
253 PASpAspGlyGlnGlnLeuGlnThrLeuPheThrThrGluGluArg 270
||||| :||||| ||||||||| ||||||| :|||
460 GGATATATGTCAGCAGCTTTGACAGGTTCTTTTACACAGAAAGAG 411
270 luArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArg 286
||||| :||||| ||||||||| ||||||| :|||
410 AAAGATCTCTGATGAGGCTAGAAAAAATTTCTAGAGAGAGCGGACA 361
287 ProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgPr 303
||||| :||||| ||||||||| ||||||| :|||
360 CCCACTGCTCTCCCTAACCTCGTGAATGAGGCTTCCCTTGAAACCGCC 311
303 OGlyTyrAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyr 320
||||| :||||| ||||||||| ||||||| :|||
310 CAACTGGAGCTTACAAACACCGCGGAGGAGGACACCTCTCTTATC 261
320 rGlnAlaIleLeuValAlaGlyLeuArgLysAlaSerArgArgProThrAsn 336
||||| :||||| ||||||||| ||||||| :|||
260 GCCGAGCTCTAGTGCAGAGCTCTCAGAGGCGC..... 230
337 LeuAlaLysValArgGluValMetGlnGlyProAsnGluProProSerVa 353
||||| :||||| ||||||||| ||||||| :|||
229 CTGGCTAGAGTAAGAGAGGCTTGCAGAGGCGAGCTGACACACCTCAGT 180
353 lPheLeuGluArgLeuMetGluAlaPheArgArgPheThrProPheAsp 370
||||| :||||| ||||||||| ||||||| :|||
179 CTTCCTTGAAGCTCTCATGAGGATATAGAGGATACACCCCTTTTACC 130
370 roThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGln 386
||||| :||||| ||||||||| ||||||| :|||
129 CCTGCTAGAGGCGCAGAAAGCTCTGATGCATGCGCTTCATTTGGTCAG 80
387 SerAlaLeuAspIleArgGlyLysLeuGlnArgLeuGlnGlyLeuGlnG 403
||||| :||||| ||||||||| ||||||| :|||
79 TCCCGTCCCGCATTAAGAAAAAGCTCAAGAGGCTGAGGCGCTCCAAGA 30
403 uAlaGluLeuArgAspLeu 409
||| :|||
29 TCATACGCTCCAGATTTA 11
seq_name: gb_est2:BG089942

```

seq_documentation_block: 645 bp mRNA EST 26-JAN-2001
 LOCUS BG089942
 DEFINITION mab86d06.k1 NCI_CGAP.SP2 Mus musculus cDNA IMAGE:3977506 3'

similar to TR:Q9WLL1 Q9WLL1 GAG PROTEIN. [1] ;, mRNA sequence.

ACCESSION BG089942
 VERSION BG089942.1 GI:12572505
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 TITLE 1 (bases 1 to 645)
 JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 Other ESTs: mab86d06.y1
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail@nih.gov
 Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

MG1:147538
 Seq primer: -40UP from Gibco
 High quality sequence stop: 422.
 Location/Qualifiers

FEATURES

source

1..645
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3977506"
 /clone_lib="NCI_CGAP.SP2"
 /lisse_type="NK cells (flow-sorted)"
 /lab_host="DH10B (T1-resistant)"
 /note="Organ: spleen; Vector: pCMV-SPORT6 (Life Technologies); mRNA made from flow-sorted NK cells, CDNA made by oligo-dT priming. Directionally cloned. Average insert size 1.5 kb. Primary library, non-amplified. CDNA library preparation: David B. Kitzman, Ph.D."
 BASE COUNT 141 a 177 c 150 g 176 t 1 others
 ORIGIN

alignment_scores:

Quality: 810.50 Length: 215
 Ratio: 4.311 Gaps: 1
 Percent Similarity: 87.442 Percent Identity: 75.349

alignment_block:

US-09-171-553B-4 x BG089942/rev ..

Align seg 1/1 to reverse of: BG089942 from: 1 to: 645

```

216 PheSerSerAlaAspLeuTyrAsnTyrPlyThrAsnHisPropheSe 232
||||| :||||| ||||||||| ||||||| :|||
642 TTTTCTTCTGTCACCTTACAAATGAAGGTC...ATCCCCCTTTT 596
232 rGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMetPheSer 249
||||| :||||| ||||||||| ||||||| :|||
595 CAGAAAACTGCAAGAGCTACCTGAGTGGTGAATCATTAATGATTTAC 546
249 lGlnProThrTyrPAspAspGlyGlnGlnLeuGlnThrLeuPheThr 265
||||| :||||| ||||||||| ||||||| :|||
545 ACCAGCCGAGCTTGGATGACCTGCCAGACCTTCTGAGACTTATTCACA 496
266 ThrGluIleArgGluArgIleLeuLeuGluAlaArgLysValProGln 282
||||| :||||| ||||||||| ||||||| :|||
495 ACCGAGAGAGAGAGAGATTTCTCTCGAGGCTCGAAAAATGTCGAGACA 446
282 yAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheP 299
||||| :||||| ||||||||| ||||||| :|||
445 CGAGGCTGGGCGCCCTGTCCAAACCTCAGCTGAGATAGTGAAGATTTC 396

```

```

299  roleuThrArgProGlyTyrAspArgAsnThrAlaGlnGlyArgGlnSer 315
395  CGCTAACCCGGCCCGCATGGATTATATACGGCATAGGTAGGGAACGA 346
316  LeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGlyAlaSer 332
345  CTGTCACAAATATGGCGGCTCTAGTGGCGGGCTTCAGATGGTGGCCG 296
332  gatgagProThrAsnLeuAlaLysValArgGluValMetGlnGlyProAsn 349
295  GCAGCCCGACAAATCTGGCCAAAGTAAGAGAGATTATGACGGACGACTG 246
349  IupProSerValAlaPheLeuGluArgLeuMetGlnAlaPheArgArgPhe 365
245  AGGCCCCCTAGCTCTCTGAAAGCGCTCATGAGAGGCTTATAGAGATAT 196
366  ThrProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAl 382
195  ACCCCATTCGACCCACGCTGCTAGGGTCMAAGGGCTCTCAGTAATATAGGC 146
382  apheIleGlyLysSerAlaLeuAspIleArgLysIleLeuGlnArgLeuG 399
145  CTTCATTGGCGCATCGTACGCTCCGACATTAGGAGAAGAAGTTACAGCGAATTG 96
399  IuGlyLeuGlnGlnAlaGlnLeuArgAspLeuValArgGluAlaGlnLys 415
95  AGGGCTTGGCGAGATTACACCAATAAAGGATGTAGTTAGACGACGAGAGAA 46
416  ValTyrTyrArgArgGluThrGlnGlnGlnGlnGlnGlnGlnGlnGlnG 430
45  GTGTATCATGGGAGAAACAGACATATAAAAGTTGAGAGAGAG 1

```

seq_name:	gb_est1:BE283560
seq_documentation block:	
LOCUS	BE283560 598 bp mRNA
DEFINITION	601102760F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3495206 5'
ACCESSION	BE283560 mRNA sequence.
VERSION	BE283560.1 GI:9159895
KEYWORDS	EST.
SOURCE	house mouse.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 598)	NIH-MGC	http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		Contact: Robert Strassberg, Ph.D.

```

FEATURES
  source
    1. 598
      cdna library Arrayed by: The I.M.A.G.E. Consortium (LNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LNL at:
      http://image.llnl.gov
      plate: LNA8545 row: g column: 15
      High quality sequence stop: 598.
      Location/Qualifiers

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="3495206"
/clone_id="NCI-CGAP_L129"
/tissue_type="spontaneous tumor, metastatic to mammary,
stem cell origin"
/lab_host="RDH10B"
/organ="lung"
/vector="pCMV-SPORT6"
/site_1="SalI"
/site_2="NotI"
/cloned_unidirectionally="yes"
/primer="Oligo dT

```

Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH

BASE COUNT	144 a	164 c	155 g	135 t
ORIGIN				

```
alignment_scores:
  Quality: 809.00      Length: 193
  Ratio: 4.649        Gaps: 0
Percent Similarity: 90.155      Percent Identity: 79.275
```

alignment_block:
MS-09-171-553R-A x BF283560

Align seg 1/1 to: BE283560 from: 1 to: 598

201 ProPrometProGlyGlnLeuGlnProLeuGlnTyrTrpProphese 217
 ||||| : : : : :
 18 CCTCCTGCCGATCCAATAGTCTACAGCCCTACAGTATTGGCCTTTTC 67

217 rSerAlaAspLeuTyrAsnTrpLysThrAsnHisProProPheSerGluA 234

234 sPpRoGlnArgLeuThrGlyLeuValGluSerLeuMetPheSerHisGln 250

118 ACCCTGCAGGACTCACTGGGTTGGTTGAATCATTAATGTATTACACACCAG 167

168 CCGACCTGGATGACTGCCAGCAGCTTCTGCAGACTCTATTGACAACCGA 217

```

267 ugiurrgcluarqllleuleuclualaarqLysasnValProclYalaa 284
|||||
218 GGAGAGAGAGAGATTCCTCCTCGAGGCTCGGAAAAACGTCGCGAGATGAGG 267
|||||

```

284 spglyargprothrnglnleuglnasnlglnleuaspmetglypheproleu 300

268 CTTGGCGCCCGCTGTCGAACATCCAGCGTAGAGATAGATGAAGGATTTCCGCTA 317
301 THRAQPIQGLYTRPASPVTYRAsNThRAIagIuGIYARGIuSerLeuHY 317

318 ACCCGGCCCGATGGATTATATACGGCATCAGGTAGGGAACGACTGTC 367

31/ siieiyaiygrinnalaleuvalalaglyleuairgyalaseraiyairg 334
:
368 CAATTATCGCCGGGTCTAGTGGCGGTCACAGAGTGTGCCCCGAC 417

334 roThrasnLeuAlaLysValArgGluValMetGlnGlyProAsnGluPro 350
|||||
|||||

351 ProSerValPheLeuGluArgLeuMetGluAlaPheArgArgPheThrPr 3677

468 CCTCAGTCTTCCTTGAAGGCTCATGGAGCTTATAGAGATATACCC 517

518 ATTCGACCCACGCTGAGGGTCAAGGCCCTCAGTAATTATGGCCITCA 567

384 leglylInSerAlaLeuAspIleArgLys 393

name: gb_gss:A2650638

documentation_block:	601 bp	DNA	GSS	1
S	AZ650638			

clone UUGC1M0520L20 R, DNA sequence.

AZ650638

CT.117965336

AZ650639 1

ORDS GSS.

```

SOURCE          mouse mouse.
ORGANISM        Mus musculus
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS         Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
                Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
                ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
                and Wright,D., Weiss,R.
TITLE           Mouse whole genome scaffolding with paired end reads from 10kb
                plasmid inserts
JOURNAL         Unpublished (2000)
COMMENT         Contact: Robert B. Weiss
                University of Utah Genome Center
                Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCC, UT
                84112, USA
                Tel.: 801 585 5606
                Fax: 801 585 7177
                Email: ddunn@genetics.utah.edu
                Insert length: 10000 Std Error: 0.00
                Plate: 0520 row: L column: 20
                Seq primer: CACACAGAAACACTTATGACC
                Class: plasmid ends
                High quality sequence stop: 601.
FEATURES
source          1..601
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUCGM052OL20"
                /clone_lib="Mouse 10kb plasmid UUCGM library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F- "
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptor DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of PWD42 (gII4732114[gb]/AFI29072.1), a copy-number
                inducible derivative of plasmid RI. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."
BASE COUNT      175 a 147 c 169 g 110 t
ORIGIN
alignment_scores:
    Quality:      800.00 Length:      199
    Ratio:        4.372   Gaps:        0
Percent Similarity: 91.960 Percent Identity: 76.382
alignment_block:
us-09-171-553B-4 x AZ650638 ..
Align seg 1/1 to: AZ650638 from: 1 to: 601
248 SerHISglNProThrTPAsPaSPCySglngInLeuGlntThrLeuPh 264
|||||.....
5 TCCTCATCAACCCACTTGGAATGTCTCAGCGCTTTGGACAGTTCCTTT 54
|||||.....
264 eThrThrgLuGLuArgGLileuLeuGLuLlaArgLySaSnValP 281
|||||.....

```

55 CACCACAGAGAAAGAGAAAGATTCGTATGATGGAGCCAGAAAAAATGTTCC 104

281 rogl1yAlaspl1yArgrProthrgInleuGlnasng11leasPmetGly 297

105 TAGAGAGAGACGGCACACCCTGCGCTCCCTTAATCTTCGGAGCGAGGCT 154

298 pheProLeuThArpProGly1YTPAsp1yrsnThra1ag1u1yArgr1 314

155 TTTCCTTGAACCGCCCACTGGAGACTACACACCGCGGAGAGTAgGGG 204

314 userLeu1yS11eYrArgrGlnAlaLeuValAlaGlyLeuArGlyAlAS 331

205 ACGCCCTCCTTGCTGTATCCGCGAGCTACAGTGGCGAGGTCTCATAGGGGCGG 254

331 eArArArProthrsnLeuAla1ySValArG1uVal1metG1ng1yPro 347

255 CTAGAGCGCCCAACCAATTTGGCTAAAGTAAAGAGG1CTTGGACGGGCGAG 304

348 Asng1uProProSerValPheLeuGlnArG1uLeuMetGln1aPheArgr 364

305 AC1GAACCAACCTCAGTCTCTCTGAGCGTCATGAGGACATATAGAG 354

364 gphenThrProPheaspProThrsSerGlnAlaGln1ySAlaSerValAla 381

355 GTACACCCCTTTTGACCCCTGCTAGAGGGGCGACAAAGCGGTGTAGCCA 404

381 eAlaPhe11eGlyInSerAlaLeuAsp11eArGly1yS1yLeuGlnArG 397

405 TGCCCTCATTTGTGTAGTCGCTCCGACATTTAAAGAAAAAGCTGCAGAG 454

398 LeuG1uGlyLeuGlnGlnAlaGlnLeuArGAspLeuValArG1uAlaG1 414

455 CTGGAGGGGCGCTCCAAAGATCATACCTCCCAAGATTGTGTAAGAACCA 504

414 u1ySVal1yTrYrArgrArG1uTrG1uG1uG1u1ySGLInArG1yS 431

505 GAAATCTATCATTAAGAGGAAACAGAAAAAGAGAGCGAGAGAGAGACA 554

431 1u1ySGLuArG1uGlnGlnGlnGlnGlnArGrAspArGrArG1ng1u 446

555 AGAAAGAAATGAGAGAGAGGAAATATAGACGGAGTGCCTGACGAG 601

seq_name: gb_AZ708792

seq_documentation_block: 820 bp DNA 24-JAN-2001

LOCUS AZ708792 GSS

DEFINITION RPI-23-238Pd_TV RPI-23 Mus musculus genomic clone RPI-23-238Pd, DNA sequence.

ACCESSION AZ708792

KEYWORDS AZ708792.1 GI:12438716

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 820)

298 pheProLeuThArpProGly1YTPAsp1yrsnThra1ag1u1yArgr1 314

155 TTTCCTTGAACCGCCCACTGGAGACTACACACCGCGGAGAGTAgGGG 204

314 userLeu1yS11eYrArgrGlnAlaLeuValAlaGlyLeuArGlyAlAS 331

205 ACGCCCTCCTTGCTGTATCCGCGAGCTACAGTGGCGAGGTCTCATAGGGGCGG 254

331 eArArArProthrsnLeuAla1ySValArG1uVal1metG1ng1yPro 347

255 CTAGAGCGCCCAACCAATTTGGCTAAAGTAAAGAGG1CTTGGACGGGCGAG 304

348 Asng1uProProSerValPheLeuGlnArG1uLeuMetGln1aPheArgr 364

305 AC1GAACCAACCTCAGTCTCTCTGAGCGTCATGAGGACATATAGAG 354

364 gphenThrProPheaspProThrsSerGlnAlaGln1ySAlaSerValAla 381

355 GTACACCCCTTTTGACCCCTGCTAGAGGGGCGACAAAGCGGTGTAGCCA 404

381 eAlaPhe11eGlyInSerAlaLeuAsp11eArGly1yS1yLeuGlnArG 397

405 TGCCCTCATTTGTGTAGTCGCTCCGACATTTAAAGAAAAAGCTGCAGAG 454

398 LeuG1uGlyLeuGlnGlnAlaGlnLeuArGAspLeuValArG1uAlaG1 414

455 CTGGAGGGGCGCTCCAAAGATCATACCTCCCAAGATTGTGTAAGAACCA 504

414 u1ySVal1yTrYrArgrArG1uTrG1uG1uG1u1ySGLInArG1yS 431

505 GAAATCTATCATTAAGAGGAAACAGAAAAAGAGAGCGAGAGAGAGACA 554

431 1u1ySGLuArG1uGlnGlnGlnGlnGlnArGrAspArGrArG1ng1u 446

555 AGAAAGAAATGAGAGAGAGGAAATATAGACGGAGTGCCTGACGAG 601

seq_name: gb_AZ708792

seq_documentation_block: 820 bp DNA 24-JAN-2001

LOCUS AZ708792 GSS

DEFINITION RPI-23-238Pd_TV RPI-23 Mus musculus genomic clone RPI-23-238Pd, DNA sequence.

ACCESSION AZ708792

KEYWORDS AZ708792.1 GI:12438716

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 820)

298 pheProLeuThArpProGly1YTPAsp1yrsnThra1ag1u1yArgr1 314

155 TTTCCTTGAACCGCCCACTGGAGACTACACACCGCGGAGAGTAgGGG 204

314 userLeu1yS11eYrArgrGlnAlaLeuValAlaGlyLeuArGlyAlAS 331

205 ACGCCCTCCTTGCTGTATCCGCGAGCTACAGTGGCGAGGTCTCATAGGGGCGG 254

331 eArArArProthrsnLeuAla1ySValArG1uVal1metG1ng1yPro 347

255 CTAGAGCGCCCAACCAATTTGGCTAAAGTAAAGAGG1CTTGGACGGGCGAG 304

348 Asng1uProProSerValPheLeuGlnArG1uLeuMetGln1aPheArgr 364

305 AC1GAACCAACCTCAGTCTCTCTGAGCGTCATGAGGACATATAGAG 354

364 gphenThrProPheaspProThrsSerGlnAlaGln1ySAlaSerValAla 381

355 GTACACCCCTTTTGACCCCTGCTAGAGGGGCGACAAAGCGGTGTAGCCA 404

381 eAlaPhe11eGlyInSerAlaLeuAsp11eArGly1yS1yLeuGlnArG 397

405 TGCCCTCATTTGTGTAGTCGCTCCGACATTTAAAGAAAAAGCTGCAGAG 454

398 LeuG1uGlyLeuGlnGlnAlaGlnLeuArGAspLeuValArG1uAlaG1 414

455 CTGGAGGGGCGCTCCAAAGATCATACCTCCCAAGATTGTGTAAGAACCA 504

414 u1ySVal1yTrYrArgrArG1uTrG1uG1uG1u1ySGLInArG1yS 431

505 GAAATCTATCATTAAGAGGAAACAGAAAAAGAGAGCGAGAGAGAGACA 554

431 1u1ySGLuArG1uGlnGlnGlnGlnGlnArGrAspArGrArG1ng1u 446

555 AGAAAGAAATGAGAGAGAGGAAATATAGACGGAGTGCCTGACGAG 601

seq_name: gb_AZ708792

seq_documentation_block: 820 bp DNA 24-JAN-2001

LOCUS AZ708792 GSS

DEFINITION RPI-23-238Pd_TV RPI-23 Mus musculus genomic clone RPI-23-238Pd, DNA sequence.

ACCESSION AZ708792

KEYWORDS AZ708792.1 GI:12438716

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 820)

298 pheProLeuThArpProGly1YTPAsp1yrsnThra1ag1u1yArgr1 314

155 TTTCCTTGAACCGCCCACTGGAGACTACACACCGCGGAGAGTAgGGG 204

314 userLeu1yS11eYrArgrGlnAlaLeuValAlaGlyLeuArGlyAlAS 331

205 ACGCCCTCCTTGCTGTATCCGCGAGCTACAGTGGCGAGGTCTCATAGGGGCGG 254

331 eArArArProthrsnLeuAla1ySValArG1uVal1metG1ng1yPro 347

255 CTAGAGCGCCCAACCAATTTGGCTAAAGTAAAGAGG1CTTGGACGGGCGAG 304

348 Asng1uProProSerValPheLeuGlnArG1uLeuMetGln1aPheArgr 364

305 AC1GAACCAACCTCAGTCTCTCTGAGCGTCATGAGGACATATAGAG 354

364 gphenThrProPheaspProThrsSerGlnAlaGln1ySAlaSerValAla 381

355 GTACACCCCTTTTGACCCCTGCTAGAGGGGCGACAAAGCGGTGTAGCCA 404

381 eAlaPhe11eGlyInSerAlaLeuAsp11eArGly1yS1yLeuGlnArG 397

405 TGCCCTCATTTGTGTAGTCGCTCCGACATTTAAAGAAAAAGCTGCAGAG 454

398 LeuG1uGlyLeuGlnGlnAlaGlnLeuArGAspLeuValArG1uAlaG1 414

455 CTGGAGGGGCGCTCCAAAGATCATACCTCCCAAGATTGTGTAAGAACCA 504

414 u1ySVal1yTrYrArgrArG1uTrG1uG1uG1u1ySGLInArG1yS 431

505 GAAATCTATCATTAAGAGGAAACAGAAAAAGAGAGCGAGAGAGAGACA 554

431 1u1ySGLuArG1uGlnGlnGlnGlnGlnArGrAspArGrArG1ng1u 446

555 AGAAAGAAATGAGAGAGAGGAAATATAGACGGAGTGCCTGACGAG 601

seq_name: gb_AZ708792

seq_documentation_block: 820 bp DNA 24-JAN-2001

LOCUS AZ708792 GSS

DEFINITION RPI-23-238Pd_TV RPI-23 Mus musculus genomic clone RPI-23-238Pd, DNA sequence.

ACCESSION AZ708792

KEYWORDS AZ708792.1 GI:12438716

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (Bases 1 to 820)

298 pheProLeuThArpProGly1YTPAsp1yrsnThra1ag1u1yArgr1 314

155 TTTCCTTGAACCGCCCACTGGAGACTACACACCGCGGAGAGTAgGGG 204

314 userLeu1yS11eYrArgrGlnAlaLeuValAlaGlyLeuArGlyAlAS 33

page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html

Plate: 238 row: p column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..820
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPI-23-238P4"
/clone_lib="RPI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 236 a 186 c 247 g 151 t
ORIGIN

alignment_scores:
Quality: 797.50 Length: 253
Ratio: 3.853 Gaps: 5
Percent Similarity: 81.818 Percent Identity: 65.217

alignment_block:
US-09-171-553B-4 x AZ708792 ..

Align seg 1/1 to: AZ708792 from: 1 to: 820

```

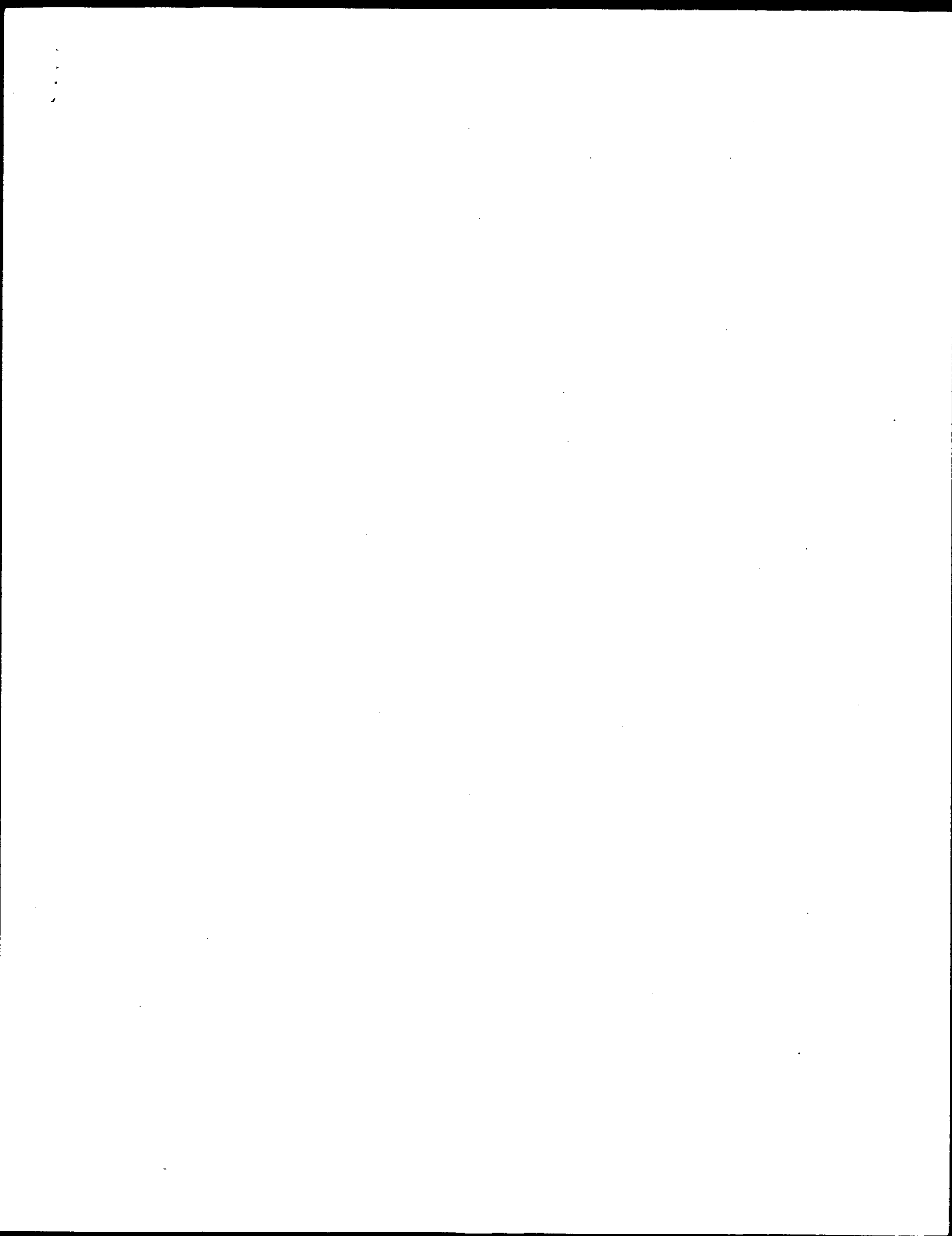
280 ValProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMe 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1  GCTTGGGGAAGAGGACTCCACCACTCTTAACCTATTGATGA 50
296 tGlyPheProLeuThrArgProGlyTyrAspTyrAsnThrAlaGlnGlyA 313
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 AGCTTTCCTTGAACCGCCCAACTGGACTACATACAGCAAGAGTA 100
313 tGlnSerLeuLysIleTyrArgGlnAlaLeuValAlaGlyLeuArgGly 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 GGGAGCGCTCTCGTATCGCCGACTAGTGGCGGGCTCAAAAGCG 150
330 AlaSerArgArgProThrAsnLeuAlaLysValArgGlnValMetGlnI 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 GCTGCACGACGGCCCTTCTTCTGCAAGGTAGAGAGGCTTGCAGGG 200
346 yProAsnGluProSerValPheLeuGluArgLeuMetGlnAlaPheA 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 ACCAGCGAGCCACCTCGCTTCTTAGAACGCGCTAATGAGAGCCTATA 250
363 tArgPheThrProPheAspProThrSerGlnAlaGlnLysAlaSerVal 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 GGAGGTATACCCCGTTCATCTTCATCGGAGGGGCAAGCAAGCAGCTGT 300
380 AlaLeuAlaPheIleGlyLysSerAlaLeuAspIleArgLysLysLeuG 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 GCTATGGCTTTCATTGGCCAGTCAGCTCCGATTTAAAAAGAAAGCTGCA 350
396 nArgLeuGlnGlyLeuGlnGlnAlaGlnLeuArgAspLeuValArgGlnA 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 AAGACTCGAGGGGCTCCAGATTTACTTTGACAGATCTAGTGAAGAGAG 400
413 lAgLysValTyrTyrArgArgGlnThrGlnGlnLysGlnGlnArg 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 CAGAGAAAGTGTTCATPAGAGAGACGGAAGATGAAAAAGAAAGA 450
430 LysGlnLysGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 446
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
451 GAGAAAAAGAGGCAAGAGACGAGAGATGACGGGACCGCTGCACAGGA 500

```

```

446 uLysAsnLeuThrLysIleLeuAlaAlaVal...GlnGlyLysSerS 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 AAGAAACCTGACAAAGATTTCTGGCCGCACTGGTAGAGAGAGACAGTTAA 550
462 eArgGlnArgGlnArgAspPheArgLysIleArgSerGlyProArgGln 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 GGAGAGATAGG..... 561
479 SerGlyAsnLeuGlyAsn.....ArgThrPr 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
562 TCAGGGAACCTGGCACTGGCGAGAGACGAGATTGCAAGACCCCAACC 611
487 o..LeuAspLysAspGlnCysAlaTyrCysLysGlnLysGlnIstPAl 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 CAGTTGAAAGAGACAGTGTGGTATTGCAAAAGAAAGGAGCATTTGGGT 661
503 aArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuG 520
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
662 TCGGGAATGCCCAAGAGAGA.....GCCAAGGTGCTGGCCCTCG 701
520 lGlnLysP 522
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
702 ATCAAGAT 709

```

Quality: 1039.00 Length: 228
Ratio: 4.833 Gaps: 0
Percent Similarity: 94.298 Percent Identity: 84.211

alignment_block:

US-09-171-553b-5 x AZ636636/rev ..

Align seg 1/1 to reverse of: AZ636636 from: 1 to: 685

```

213 AsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPr 229
|||||
685 AAFACACCACTTCTGCCGTTGAAAACAGGAGGACCAATGACTATCGGCC 636
229 OValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisPro 246
|||||
635 GGTACAGAGCTCCGGGAGGTAAACAAAAGGCTCTGAACANTCACCCCA 586
246 hValProAsnProTyrAsnLeuLeuSerAlaLeuProProGluArgAsn 262
|||||
585 CAGTCCCGAACCCGTACATTTATTAACCTCTCTCCACCTGAGAGAACCA 536
263 TTPYrThValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHi 279
|||||
535 TGTATACAGTCTGTGACTTAAAGATCCCTCTTTGCGCTGCTTCA 486
279 sProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrG 296
|||||
485 CCTTAAAGATCAGCTCCTGTTGCTTTGAAATGAGAGGAGCCAGAGGCG 436
296 lYArgThGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsn 312
|||||
435 GACACAGCTGCACTACTGACTGAGCTACACAGAGGAGTTCAAAAT 386
313 SerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheAr 329
|||||
385 TCCTCCACCTGTTGACGAGGCGCTCCATCGGATCTTGGCGCTTTGCG 336
329 gIleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeu 346
|||||
335 CGCTCGAAACCTCAGCTTACCTACTACATGATGATGATGATCTCTGG 286
346 euAlaGlyAlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeu 362
|||||
285 TTGTGGGCGCTCGAAGAGAGCTGTGCACAGGAGGAGCTGAGAGCTCTC 236
363 LeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnI 379
|||||
235 GCAGAACTGAGTGAAGTGGGGTATCGAGTTTCGGCTAAAAAGCACAAAT 186
379 eCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgLysGlnA 396
|||||
185 CTGTCAAACTGAGTGAAGTGGGGTATCGAGTTTCGGCTAAAAAGCACAAAT 136
396 rGTrPLeuThrGluAlaArgLysLysThrValGlnIleProAlaPro 412
|||||
135 GATGCTCAGCAGAGGCGCGAAGAACTGTTATGATGATGCCATGCCCA 86
413 ThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysAr 429
|||||
85 ACTACCCAGGAGGAGTGAAGTGGGGTATCGAGTTTCGGCTAAAAAGCACAAAT 36
429 gLeuTrpIleProGlyPheAlaThrLeuAlaAla 440
|||||
35 ACTCTGATTCAGAGCTTTCACAACTGACAGCA 2

```

seq_name: gb_gss: BH090616

seq_documentation_block:

LOCUS BH090616 791 bp DNA GSS 18-JUL-2001
DEFINITION RPI-24-356L4.TJ RPI-24 Mus musculus genomic clone RPI-24-356L4,
DNA sequence.
ACCESSION BH090616
VERSION BH090616.1 GI:14910521

KEYWORDS

GSS.
house mouse.

SOURCE

Mus musculus

ORGANISM

REFERENCE

1 (bases 1 to 791)

AUTHORS

Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akintore, B., Levins, M., Russell, D., de Jong, P., and Fraser, C.M.

TITLE

Mouse BAC End Sequences from Library RPI-24

JOURNAL

Unpublished (1999)

COMMENT

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@ligr.org

Clones are derived from the mouse BAC library RPI-24. For BAC library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end

plate: 356 row: L column: 4

Seq primer: SP6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..791

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPI-24-356L4"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTRABAC1; Site-1: BamHI; Site-2: BamHI; RPI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 216 a 201 c 188 g 186 t
ORIGIN

alignment_scores:

Quality: 1011.50 Length: 257
Ratio: 4.341 Gaps: 2
Percent Similarity: 90.661 Percent Identity: 73.541

alignment_block:

US-09-171-553b-5 x BH090616 ..

Align seg 1/1 to: BH090616 from: 1 to: 791

```

118 ValLeuThrLeuGlnLeuAspAspLutYrArgLeuTyrSerProGlnVa 134
|||||
25 GTCTTAGCCCTCAGCTGAGGAGAAATACCATGATGATGAA...GGCAT 71
134 lYsProAspGln.....AspIleGlnSerPheLeuGlnGlnPheProG 149
|||||
72 AAAGCAACAGAGAGTGTACCTTAAAGAGCTGAGCTGCTTCCCTTA 121
149 lAlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProPro 165
|||||
122 GAGCATGGGAGAGAGCTGAGGAGATGAGATGAGTGCAGAGTCCCT 171
166 GlnValIleGlnLeuLysAlaSerAlaThrProValSerValArgGly 182
|||||
172 GTTTTGTGGGAGCTGAGAGATGCAACCCCTATAGAGAGTGCACATA 221
182 rProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgL 199
|||||
222 TCATATGAGCCGTGAGGAGAGATGATGATGAGGAGCCATATTCAGAGAC 271

```

199 euilelenginglylleuValProvalGlnSerProtrpAsnThrPro 215
 272 TATTAACAATTAGTATTGTTGACCTGCGACGACCTGGAATACCTCT 321
 216 leuLeuProvalArglySerProglyThrAsnAspTyrArgProvalGlnAs 232
 322 TTTACTCCGGTAAAGAACCTGGCAGAGTGAATACCGACCGATCCAA 371
 232 pleuArgGluValAsnLysArgValGlnAspIleHisProThrValProA 249
 372 CCGAAGAGAGGTCAATAGAGGTCACAGATATTCACCACTGTTCCTA 421
 249 snpProTyrAsnLeuLeuSerAlaLeuProProGlnArgAsnTrpTyrThr 265
 422 ATCCCTATACCTCCTCAGCTCACTCCCTCCAGAGGAAATGATACACA 471
 266 ValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThSe 282
 472 GTACTTGATTAAAGATGCTTCTTTGCTGGAAGTTACATCCCTCCAG 521
 282 rginProLeuPheAlaPheGluTyrPArgAspProGlyThrGlyArgThrg 299
 522 CCAAGCCCATATTCGCTTTGAGTGGAGAGATCCTGACACTGGACAAATGG 571
 299 lylGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 315
 572 GACAAATTGACCTGGACACGCTTACCCAGAGATTCAAAAACCTCCGCCAC 621
 316 llePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHis 332
 622 CTTTGTGATGAGGCTCTACATGAGATTTAGCTCTCTCTAGCGAANA 671
 332 sProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyA 349
 672 CTCACAGGTAAACCTGCTCTCATATGTGTGATGACTTCTCTGCTGC.A 720
 349 lathLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuLeuLeu 365
 721 CCACAGAGAGAGATGTTGGCAGGAGACCAAGAGATGCTAGTACGACTG 770
 366 SerAspLeuGlyTyrArgAla 372
 771 GGTGAGTTGGTTACCGGGCC 791
 seq_name: gb_est2:BI107577
 seq_documentation_block:
 LOCUS BI107577 799 bp mRNA EST 26-JUN-2001
 DEFINITION 602894023F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5039044 5',
 mRNA sequence.
 ACCESSION BI107577
 VERSION BI107577.1 GI:14558470
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM1107 row: n column: 05
 High quality sequence stop: 743.

FEATURES

source

Location/Qualifiers
 1..799
 /organism="Mus musculus"
 /strain="C57BL/6J (f1er1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5039044"
 /tissue="NCI_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 211 a 215 c 218 g 154 t 1 others
 ORIGIN

alignment_scores:

Quality: 1004.00 Length: 263
 Ratio: 4.236 Gaps: 5
 Percent Similarity: 90.114 Percent Identity: 75.285

alignment_block:

US-09-171-553B-5 x BI107577 ..

Align seg 1/1 to: BI107577 from: 1 to: 799

507 ArgArgProValAlaTyrLeuSerLysLysLeuAspProValAlaSerGI 523
 20 CGGGCCCGCGTAGCGCTGTTATCAAGAAACTGGACGCTGTGGCCAGTGG 69
 523 yTTPProValGlyLeuLysAlaIleAlaIleAlaValAlaIleValLysA 540
 70 ATGGCCCTCCTCGCTCGGAGGATAGCAGCCAGCGCTGTGTGTAGTAAG 119
 540 spAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHis 556
 120 ATGTGACAAACTGACATATGCGCAGAAATGTACTATATAGTGGCCACAC 169
 557 AlaLeuGlnAsnIleValArgGlnProProAspArgTrpMetThrAsnAl 573
 170 TCTCTGAGAGCATCATCAGACCAACCGACCGCTGATGACCAACGC 219
 573 ArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheA 590
 220 CCGATGACGCGACTACCAAGCCCTATTGCTGACAGACGAGTAAGTTTG 269
 590 lAProPAlaAlaLeuAsnProAlaThrLeuLeuProGlnGluThrAsp 606
 270 CACCCCGACCATTTCTCAACCCCGCTCTACTACT...GAGGCTGAC 316
 607 GluProValThrHisAspCysHisGlnLeuLeuIleGlnGluThrGlyA 623
 317 GAGGCCCTGCACATAGTGTGAAGAAATCTGGCAGAGACAGCTGGAAT 366
 623 lArgLysAspLeuThrAspIleProLeuThrGlyValLeuThrTrp 640
 367 CCGGCGACCTCCACAGACCACTTGGCCAGG...GCCATGACTTGGT 413
 640 heThrAspGlySerSerTyrValValGlnGluLysArgMetAlaGlyAla 656
 414 TCACGAGGAGAGAGAGCTTGTGTGTAGAAAGTAAGCGGAGGCTGGGCT 463
 464 GCAGTAGTAGTGAACAGCGCTGTCAATATGCGCAGACAGCTCCGAGAG 513
 657 AlaValValAspGlyThr.ArgThrIleTyrPalAspSerLeuProGluG 673
 673 lYThrSerAlaGlnLysAlaGlnLeuMetAlaLeuThrGlnAlaLysArg 689
 514 GTACATAGCTCAAAAAGCGGAACTATATGCAATTAATCAAGCCTTAAGG 563
 690 leuAlaGlnGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPh 706

678 salagileuMetalaurhnglnAlaLeuArgLeuAlaGlnGlyIys 695
 |||||
 780 GGCTAGCTCCTGAGCTAAAGCTCCGCGCAAGAGTGGC 828
 695 erlleAsnIleTyrThrSperArgTyrAlaPhe 706
 |||||
 829 A.ATCGAATATTTACACTGACAGCCATGCTTT 863

seq_name: gb_gss:AZ980516

seq_documentation_block: 682 bp DNA GSS 27-APR-2001
 LOCUS AZ980516 2M0257006R Mouse 10kb plasmid U0GC2M library Mus musculus genomic
 DEFINITION clone U0GC2M0257006 R, DNA sequence.
 ACCESSION AZ980516
 VERSION AZ980516.1 GI:13851743
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 682)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0257 row: 0 column: 06
 Seq primer: CACACAGGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 682.
 Location/Qualifiers
 1..682
 source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="U0GC2M0257006"
 /clone_lib="Mouse 10kb plasmid U0GC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv: Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (G147321149b|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 143 a 171 c 189 g 179 t

ORIGIN

alignment_scores:
 Quality: 998.00 Length: 227
 Ratio: 4.752 Gaps: 1
 Percent Similarity: 92.511 Percent Identity: 81.938

alignment_block:
 US-09-171-553b-5 x AZ980516/rev ..

Align seg 1/1 to reverse of: AZ980516 from: 1 to: 682

179 ValArgGlnTyr.ProLeuSerArgGlnAlaArgGlnGlyIleTyrPro 195
 |||||
 681 GTGAGCAATATACCCCATGAGCAAGAGAGTACGAGGCGCATCCGGCC 632
 195 IsValGlnArgLeuIleGlnGlnGlyTleuValProValGlnSerPro 211
 |||||
 631 ATATCCAGAGCTTGCTAGACCAAGAGAGTTTATGCGCTGTCAATCCCC 582
 212 TTPAsnThrProLeuProValArgLysProGlyThrAsnAspTyrAr 228
 |||||
 581 TGGAAATACACCACTTCTGCGGTTGCAAAACGAGGACCAATGACTATG 532
 228 gProValGlnAspLeuArgGlnValAsnLysArgValGlnAspIleHisP 245
 |||||
 531 CCCGGTGCAAGAGCTCCGGGAAGTTAACAAAGAGGTCCTGGACATTCACC 482
 245 roThrValProAsnProTyrAsnLeuLeuSerAlaLeuProGluArg 261
 |||||
 481 CCACAGTCCCGAACCCTACAAATTTATTAAGCTCTCCACCCGAGAGA 432
 262 AsnTPTyThrValLeuAspLeuLysAspAlaPhePheCysLeuArg 278
 |||||
 431 ACATGTATACACTCTGAGACTTAAGATGCTTTTGTGCTGGCTT 382
 278 uHisProThrSerGlnProLeuPheAlaPheGlnTyrArgAspProGlyT 295
 |||||
 381 GCACCTTAGAGTACAGCTCTGTTGCTTGTGAATGAGAACCAACCGAG 332
 295 hngIyArgThrGlnGlnLeuThrTPTThrArgLeuProGlnGlyPheLys 311
 |||||
 331 GCGGAGAGAGCTGTCACTAACCTGAGCTAGAGCTACACAGGGGTTCAAA 282
 312 AsnSerProThrIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnH 328
 |||||
 281 AATTCCTCCACCTGTTTACAGAGACCTCCATCGGATCTGCGCTTT 232
 231 TCGCGCTCGAAACCTCAGACTTACCTACTACAGTATGATGATCTCT 182
 345 euleuAlaGlnAlaThrLysGlnAspCysLeuGlnGlyThrLysAlaLeu 361
 |||||
 181 TGGTCCGCGACCTCGAAGAGAGCTGTGCACACAGGAAGTACAGAGCTC 132
 362 LeuLeuGlnLeuSerAspLeuGlyTyrArgAlaSerAlaLysAlaGln 378
 |||||
 131 CTCACAGAACTGAGTACTTGGGGTATCGAGTTTGGCTAAAGGACACA 82
 378 nIleCysArgArgGlnValThrTyrLeuGlnGlyTyrSerLeuArgGlyGly 395
 |||||
 81 AATTGTCTCAAACTGAGGTAACTTACCTGGGGTATACCTCGAGGGGGTA 32
 395 InArgTyrLeuThrGlnAlaArgLysLys 404
 |||||
 31 AAAGATGCTCACAGAAAGCCGGAAGAG 3

seq_name: gb_est2:BI152889

seq_documentation_block: 813 bp mRNA EST 05-JUL-2001
 LOCUS BI152889 602918048F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5068643 5',
 DEFINITION mRNA sequence.

ACCESSION B1152889
 VERSION B1152889.1 GI:14612890
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 813)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Plate: L14M1184 row: 0 column: 12
 Location/Qualifiers
 1. 813
 /organism="Mus musculus"
 /strain="C57BL/6J (f1er1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5068643"
 /clone_lib="NCI CGAP Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary. Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: PCMV-SPORE6; Site: 1; Salt; Site: 2; Note: Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
 BASE COUNT 225 a 218 c 205 g 165 t
 ORIGIN
 alignment_scores:
 Quality: 993.00 Length: 261
 Ratio: 4.208 Gaps: 4
 Percent Similarity: 90.421 Percent Identity: 76.628
 alignment_block:
 US-09-171-553b-5 x B1152889 ..
 Align seg 1/1 to: B1152889 from: 1 to: 813
 150 AATPAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProG1 166
 |||||
 2 GCGTGGCGAGAGCGGAGAGTGGGATGGCAAAAGTG.GTCCCCCTGT 50
 |||||
 166 nValIleGlnLeuLysAlaSerAlaThrProValSerValArgLntyrP 183
 |||||
 51 GGTGTTGAACCTTAAGTCGGGGCCACCCCTATGAGGGGTCCGAAATATC 100
 |||||
 183 rOlSerArgGluAlaArgGluGlyIleThrPro.HisValGlnArgLe 199
 |||||
 101 CCAATGAGAGAGAGCTCAAGGGTATACGCCCTCCAAATCAACAACACT 150
 |||||
 199 uIleGlnGlnIleLeuValProValGlnSerProThrPasnThrProL 216
 |||||
 151 GCTCAACAAGGATTTTGTTCATCAAAATCCCTTGAACACTCTCTC 200
 |||||
 216 euleProValArgLysProGlyThrAsnAspTyrArgProValGlnasp 232
 |||||
 201 TACTTCCAGTAATAAAACCAAGGACGAGGACTATCTCCAGTACAGGAC 250
 |||||
 232 LeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAs 249
 |||||
 251 CTTAGAGAGTCAACAAGAGAGTTCAGAGCATATACACCCACGGTGCCAA 300

249 nProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTyrThrV 266
 |||||
 301 TCCTTATAACCTCCTCAGACACTTGGCACCCTGGTCGACATGTGTACAG 350
 |||||
 266 alleuAspLeuLysAspAlaPhePheCysLeuArgLeuHisProThrSer 282
 |||||
 351 TCCTGATTCACAAAGAGCGCTTTTGTGTGAGTTTACACCCCAACAC 400
 |||||
 283 GlnProLeuPheAlaPheGlnTyrPargAspProGlyThrGlyArgThrG 299
 |||||
 401 CAGCCCTGTGCTTCCCTTTCGATGCGAGACTCCGAGAGTGGACACCG 450
 |||||
 299 yGlnLeuThrTyrThrArgLeuProGlnGlyPheLysAsnSerProThrT 316
 |||||
 451 ACAGCTCACATGAGACGAGGCTGCTCAGGATTCAGAACTCGCCACTT 500
 |||||
 316 lPheAspGluAlaLeuHisArgAspLeuAlaAsnThrArgIleGln.Hi 332
 |||||
 501 TGTTCGATGAAGCCCTACACCGAGATCTGCTTTTTCGAGCCAAATTA 550
 |||||
 332 sProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyA 349
 |||||
 551 CCCACAGTGTACTCTTCTGCAATATGTAGATGCTGCTCTGCTGACG 600
 |||||
 349 lAThrLysGlnAspCysLeuGlnGlyThrLysAlaLeuLeuGlnLeu 365
 |||||
 601 AAACACGCGAGACTGTGAATTTGGGACCAAAACCTCTGGGCGAGTT 650
 |||||
 366 Ser.AspleuGlyTyrArgAlaSerAla.LysLysAlaGlnIleLysAr 381
 |||||
 651 AGGTAAAGCCCTGGGTATCGGGCTCTGCTTAAAGGCTCAGTATAGCCA 700
 |||||
 381 gArgGluValThrTyrLeuGlnTyrSerLeuArgGlyGlnArgTyrP 398
 |||||
 701 GATAGAGTGTACTTACTTATGATATGCTTTGAGAGATGAGACCAAGTGTAG 750
 |||||
 398 eutThrGluAlaArgLysLysThrVal 406
 |||||
 751 TCACAGAGCCAGACAAAGACTGTA 776
 |||||
 seq_name: gb_gss: A2627566
 seq_documentation_block:
 LOCUS A2627566 690 bp DNA GSS 13-DEC-2000
 DEFINITION LM0469L20F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 clone UNGC1M0469L20 F, DNA sequence.
 ACCESSION A2627566
 VERSION A2627566.1 GI:11749756
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0469 row: L column: 20
 Seq primer: CTTGTAAACGACGGCCAGT

Class: plasmid ends
High quality sequence stop: 690.
Location/Qualifiers

1. 690
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="U08C1M0469L20"
/clone.lib="Mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114gb/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 182 a 192 c 165 g 151 t
ORIGIN

alignment_scores:
Quality: 988.00 Length: 228
Ratio: 4.682 Gaps: 1
Percent Similarity: 92.544 Percent Identity: 81.579

alignment_block:
US-09-171-553B-5 x AZ627566 ..

Align seg 1/1 to: AZ627566 from: 1 to: 690

243 ILEHSPROTHVALPROASNPROTYRASNLEULEUSERALALEUPROPR 259
|||||
7 ATACACCCACAGTCCCAATCTTATACCTCCACAGCCTGGCCACC 56
259 OGLHARGASNTPTTyrThrValLeuAspLeuLysAspAlaPhePheCysL 276
I |||||
57 TGGTGGACATGGGTACACAGTCTGATCTCAAGATGCTTTTCTGTT 106
276 euARLeuHisProThrsertInProLeuPheAlaPheGluTTPARGASP 292
|||||
107 TGAGGTACACCCCAACAGCCAGCCCTTGTCTTGCATGATGGCAGAC 156
293 PROGLYTHNGLYARGTThrGlyLeuThrTrpThrArgLeuProGlnG 309
|||||
157 TCCGAGATGGACAGCCGACAGCTTCAATGAGAGGAGGCTGCTCAGGG 206
309 yPheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeu 326
|||||
207 ATTCAAGAACTCGCCACTTTGTCGATGAAGCCCTACACGAGATCTGT 256
326 IASpPheArgIleGlnHisProGlnValThrLeuGlnTyrValASP 342
|||||
257 CTCTTCTCCGAGCCCAATTAACACAGGAGTCTTTCATATAGTATAT 306
343 ASPLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLy 359
|||||
307 GATCTGCTCTTACGCTGCAGAAACAGCAGAGACTGTGAATATGGACCCA 356
359 sAlaLeuLeuGlnLeuSerAspLeuGlyTyrArgAlaSerAlaLysL 376

357 AAACCTCCTGGGCGAGTTAGTAGCTGGGGTATCGGCGCTGCTGCTAAA 406
376 ySAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArg 392
|||||
407 AGGCTGAGTATATGCGACAGTGAAGTGAACCTACCTAGATATGCTTGAGA 456
393 GLYGLNArgTrpPLeuThrGluAlaArgLysLysThrValValGlnI 409
|||||
457 GATGACACAGCGTGGCTCAGACAGAGCCAGAAACACACTGTTATGCAGAT 506
409 ePrioAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaG 426
|||||
507 CCGGCGCCCAACACACTGCTCGCCAGGTAAAGAGATCTCTGGGACGCGCG 556
426 lYpHeCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeu 442
557 GGTTCACAGACTCTGATTCGCGGATTCGCCACAGCGCAGCTCCCTTA 606
443 TyrProLeuThrLysGluLysGlyGly.PheSerTrpAlaProGluHisG 459
|||||
607 TATCCTTACACACAGAGAGAGAGGAGATTCACCTGACACAGACATC 656
459 lNysAlaPheAspAlaIleLysLysAlaLeu 469
657 AGCTAGCCTTTGAACATCTCAAAAAGGCACTG 688

seq_name: gb_gss:AZ717731

seq_documentation_block:

LOCUS AZ717731 817 bp DNA GSS 24-JAN-2001
DEFINITION RPI-24-15511.TJ RPI-24 Mus musculus genomic clone RPI-24-15511,
DNA sequence.

ACCESSION AZ717731 GI:12456936

VERSION AZ717731.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other_GSSs: RPI-24-15511.TV

Unpublished (1998)

Mouse BAC End Sequences from Library RPI-24

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@ligr.org

library availability, please contact Pieter de Jong

(pdejong@mail.chc.org). Clones may be purchased from BACPAC

Resources (http://www.chori.org/bacpac/ordering.htm). BAC end

plate: 155 row: I column: 1

Seq primer: SP6

Class: BAC ends.

FEATURES
Location/Qualifiers

1. 817
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="RPI-24-15511"
/clone.lib="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARAC1; Site_1: BamHI; Site_2: BamHI;
RPI-24 Mouse BAC Library produced by Pieter de Jong. The

Library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 243 a 209 c 204 g 161 t
ORIGIN

alignment_scores:
Quality: 981.00 Length: 276
Ratio: 4.105 Gaps: 3
Percent Similarity: 86.594 Percent Identity: 68.841

alignment_block:

US-09-171-553B-5 x AZ717731 ..

Align seg 1/1 to: AZ717731 from: 1 to: 817

```

534 ValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnI 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 GTAGCCCTGTGATTAAAGATGCTGACCAATTGACCAATGGGACAAACAGGT 52
550 eThValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProProA 567
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 GACTGTGTGGCCCTCATGCTTAGAAAGTACGGGGCTGACCACTG 102
567 sPArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeu 583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 ACAGATGATGACAAATGCCGAGATGCCACTATCAGAGCGCTGCTGTA 152
584 ThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrIle 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 AATGACCGGTAACTTTCGCCCTGACCTGACCTCAACCCAGCTAACCT 202
600 uLeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeu 617
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 TCTCCCTCAACAAATGATTCGCTCCGACATCATATGATGACATCC 252
617 euIleGluIleThrGlyValAlaArgLysAspLeuThrAspIleProLeuThr 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 TCGCTGAGAAACTGGGACAGAGTACCTGACTGACCAACCCCTGGCCT 302
634 GlyGluValLeuThrTyrPheThrAspGlySerSerTyrValValGlu 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 GGA...GCTCCAGTTGGTACAGGACGCGACAGTTCCTGATAGAGGG 349
650 yLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIleTrpA 667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 AAAGCAAAAGCGTGAAGCTCGGCTGATGATGGCAAAAGATTTGGG 399
667 laSerSerLeuProGluGluThrSerAlaGlnLysAlaGluLeuMetAla 683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 CAAGCGCTTGGCTGAGAGACGTCGGGCACAAAGCGTGAACCTGATAGCG 449
684 leuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrTh 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 CTTATTCAAGCCCTCCGAGAGGCTAAAGGTAGATCATTAACATCTACAC 499
700 rAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrL 717
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 TGACAGCCGCTATGCTTTGGTACAGCACATCCATGGGCGCATCTACA 549
717 yGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlu 733
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 GGCAGCAAGGCTATTGACCTCGCTGCTAAAGACATTAAAAACAAGAA 599
734 GluIleLeuSerLeuGluAlaLeuHisLeuProLysArgLeuAlaI 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 GAATTTCTGCGCTGTGGAGCCATACATGCACTTAAGAGTGGCCAT 649
750 eIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyA 767
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
650 CATCCACTGCCCGCGCCACCAAGAGAGACAACTTGTGTGGCAGGGGA 699

```

seq_name: gb_gss:A2715956

seq_documentation_block:

LOCUS A2715956 789 bp DNA GSS 24-JAN-2001
DEFINITION RPCI-24-155A1.TJ RPCI-24 Mus musculus genomic clone RPCI-24-155A1,
DNA sequence.
ACCESSION A2715956
VERSION A2715956.1 GI:12453181
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 789)
REFERENCE Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinfet,B., Levins,M.,
Tsegaye,G., Geer,K., Kroll,M., Shvartsbeyan,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
OTHER GSSs: RPCI-24-155A1.TV
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdjong@email.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderframe.htm>). BAC end
page: http://www.tigr.org/cdb/bac-ends/mouse/bac_end_intro.html
Plate: 155 row: A column: 1
Seq primer: SP6
Class: BAC ends.

FEATURES

source

1..789
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-155A1"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; site_1: BamHI; site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 230 a 200 c 195 g 164 t
ORIGIN

alignment_scores:
Quality: 972.50 Length: 253
Ratio: 4.228 Gaps: 1
Percent Similarity: 90.909 Percent Identity: 73.123

alignment_block:

US-09-171-553B-5 x A2715956 ..


```

303 TTGGCAGCTGGTGGACATGTACACAGTCTGTGATCTCAAGACGCTTT 352
273 ePhcysLeuArgLeuHisProthrSerGlnProleuPheAlaPheGluT 290
353 TTTCTTTTGGAGTTACACCCACAGCCACGACCCCTTTGCTTTCGAAAT 402
290 rPArsgAspProglYThrGlyArgThrGlyGlnLeuThrTrpThrArgLeu 306
403 GGGAGAGCTCCGAGAGTGGACAAAGCCGAGACGCTCATGAGCAGAGGCTG 452
307 ProGlnGlyPheLysAsnSerProthrTrpIlePheAspGluAlaLeuHisAr 323
453 CCTCAGGAGATTCAAGAACTCGCCACTTTGTTGATGAAGCCCTACACCG 502
323 gAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnT 340
503 AGATCTTGTCTTTTCCGAGCAATTAACCCACAGGAGTCTTTCGCAAT 552
340 yTValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGlu 356
553 ATGTAGTACCTGCTCTAGCTGCAAGAAACACGAGAGACTGTGAATTT 602
357 GYTThrLysAlaLeuLeuLeuLeuLeuSerAspLeuGlyTyrArgAlaSe 373
603 GGGACCCAAACTCTCGGGCGAGTTAGTTAAAGCTGGGATACGGGCTTC 652
373 rAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrS 390
653 TGCTAA.AAGGCTCACTTATGCCAGATAGAACTGACCTACCTAGATATG 701
390 eTLeuArgGly GlyGlnArgTyrPleuThrGluAlaArgLysLysThrVa 406
702 TCTTGAGCAGATGGCGCCAGCGGTG.....TCCAGAAACAGAAAAACAG 745
406 lValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeuG 423
746 TGTTCAGATCCCGAGCCCAATGTGCCAGTA.....GAGTCTCTGG 789
423 lYThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAla 439
790 GGCCCGCGGGT...TGCGACCTG.....TGCGACCTG..... 809
440 AlaProLeuTyrProLeuThrLysGluLysGlyLysPheSerTrpAlaPr 456
810 ATCCGATTTGCCA.....TGGGCTGC 832
456 oGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaP 473
833 TTGTTCCCTACCAAGAGGATCCGCCCAAGACACTGCTTGAACCAAA 882
473 roAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValasp 489
883 GCGGGGCGCGGGGCCCA...GTAACTCCCT..... 911
490 GluArgGlyGlyValAlaArgGlyValLeuThrGlnThrLeuGly 504
912 CGGGAAGAAGGGGGCCAGCGGAGACGACCAAAAAAGGGGC 956

```

seq_name: gb_gss:BH106844

seg_documentation_block:

LOCUS BH106844 842 bp DNA GSS 19-JUL-2001
 DEFINITION RPECI-24-352A17.TVB RPECI-24 Mus musculus genomic clone
 RPECI-24-352A17, DNA sequence.

ACCESSION BH106844
 VERSION BH106844.1 GI:14937281

KEYWORDS GSS.

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 842)

REFERENCE

1 (bases 1 to 842)

AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tseaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPECI-24
 Unpublished (1999)
 Other_GSSs: RPECI-24-352A17.TVB
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPECI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@tigr.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
 Plate: 352 row: A column: 17
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..842
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPECI-24-352A17"
 /clone_lib="RPECI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTABBAC1; Site_1: BamHI; Site_2: BamHI;
 RPECI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTABBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 221 a 228 c 200 g 193 t
 ORIGIN

alignment_scores:

Quality: 966.50 Length: 267
 Ratio: 4.221 Gaps: 1
 Percent Similarity: 85.768 Percent Identity: 70.412

alignment_block:

US-09-171-553B-5 x BH106844 ..

Align seg 1/1 to: BH106844 from: 1 to: 842

```

373 SerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 389
      ::::::::::::::::::::::::::::
18 ACCACTAAGAAAGCTCAGTATGCCATTTGCACTGACCTAGGATTA 67
389 rSerLeuArgGlyGlyGlnArgTyrPleuThrGluAlaArgLysLysThr 406
      ::::::::::::::::::::::::::::::
68 CTCCTCTCGAAGAAAGCAGTGTTCACCGAAGCTCGAAAGAAAGACTG 117
406 alValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeu 422
      ::::::::::::::::::::::::::::::
118 TGATGCAAGATCCCAACCCCAACCACTGTGCAAGATAGCTGATTTCTG 167
423 GlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAl 439
      ::::::::::::::::::::::::::::::
168 GGAACGCGCAGGCTTTGTAGACTGTGAATFACGAGGTCACATGTGGG 217
439 aaAlaProLeuTyrProLeuThrLysGluLysGlyLysPheSerTrpAlaP 456
      ::::::::::::::::::::::::::::::
218 GGCACCTGTGACCACTCACTAAAGAAAGTCCGCTTACCTGACGACG 267
456 roGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
      ::::::::::::::::::::::::::::::
268 AGGAGACGCAAAAGCCTTAAACAGCATAAAGGCTGCTGCTTGCAGCC 317
473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValas 489

```

```

|||||
318 CAGCGCTGGCCCTGCCAGATTGACTTAACCTTTGCCCTATACCTGTA 367
489 pgluarGLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProT 506
|||||
368 TGACCACGCTGGAGTAGCCCGCAGAGTCTCTGACTCAGACTCTGGGGCC 417
506 rpatGATGProValAlaTyrLeuSerLysLysLeuAspProValAlaSer 522
|||||
418 GGAAGAGGCCAGTGGCTTACTGCTCTAATAAAATTTGAGCCAGTTGCCAGT 467
523 GlyThrProValLysLeuLysAlaIleAlaValAlaIleLeuValLys 539
|||||
468 GAATGGCCCTCTCTCTCTAATAAGCATGCTGTATACCTCTCTGTGTA 517
539 saspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProH 556
|||||
518 AGATACCTGGCAAGTTACTCTAGCTCAGCATGTGAGTTTATATCCTCC 567
556 ISAlaLeuGlnAsnIleValAlaArgInProProAspArgTMetThrAsn 572
|||||
568 ATGCGTTAAAGATATTGTATGACAGCCCTGACCATGTGATGACAAAT 617
573 AlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGlnArgValThrP 589
|||||
618 GCCCGATGACCCACATACAGAGCTGTGTATATGACGAGTACTTT 667
589 ealAProProAlaAlaLeuAsnProAlaThrLeuLeuProGlnGluThra 606
|||||
668 TGCCCTACTGTCATCCACATCTACACCCCTCTCCCAAGAGGATG 717
606 spGIuProValThrHisAspGlyHisGlnLeuLeuIleGlnGluThrGly 622
|||||
718 CCTTACTGCTGGTACCAATGCTGTATTTCTGCTGGAAGAATGGA 767
623 ValArgLysAspLeuThrAspIleProLeuThrGlnGluValLeuThrTr 639
|||||
768 TCGAAGAAAGATCTGACTGAC.....CGAAGCTTG 796
639 P 639
797 G 797

seq_name: gb-955:A2715272
seq_documentation_block: 755 bp DNA 24-JAN-2001
LOCUS A2715272
DEFINITION RPI-24-154N5.tj RPI-24 Mus musculus genomic clone RPI-24-154N5,
DNA sequence.
ACCESSION A2715272
VERSION A2715272.1 GI:12451823
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 755)
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPI-24-154N5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC

```

```

FEATURES
Source Location/Qualifiers
1..755
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPI-24-154N5"
/clone_id="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1, Site_1: BamHI; Site_2: BamHI;
RPI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MhoI partially digested male C57BL/6J
DNA."

BASE COUNT 202 a 194 c 190 g 169 t
ORIGIN

alignment_scores:
Quality: 961.00 Length: 253
Ratio: 4.271 Gaps: 3
Percent Similarity: 88.933 Percent Identity: 75.099

alignment_block:
US-09-171-553b-5 x A2715272 ..

Align seg 1/1 to: A2715272 from: 1 to: 755

273 PhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheG 289
|||||
5 TTTTCTGTTGGGT...TCACCCACAGCCAGCCCTGTTCCTTCCA 51
289 uTrParGAspProGlyThrGlnArgThrGlnLeuThrTrpThrArgL 306
|||||
52 ATGGTGAGACTCCGAGAGTGAGACAGCGGACAGCATGAGCAGAGGC 101
306 euProGlnGlyPheLysAsnSerProThrIlePheAspGlnAlaLeuHis 322
|||||
102 TGCCCTCAGGATTCAGAAACCTCGCCCATTTGTTCATGTAAGCCCTAC 151
323 ArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuGln 339
|||||
152 CGAGATCTTGCTCTTTCCAGCCAAATTAACCCAGAGTGATTTCTTCCA 201
202 ATATGTAGATGACCTGCTCTAGCTCAGAAACAGCAGAGACTGTGAAA 251
339 nTyrValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeu 356
|||||
356 LuGlyThrLysAlaLeuLeuLeuGlnLeuSerLeuGlyTyrArgAla 372
|||||
252 TTGGGACCCAAACCTCTGGGAGTTAGGTAAGCTGGGGTATCGGGCC 301
373 SerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGly 389
|||||
302 TCTCTCTAATAAAGCTCAGTATAGCCAGATTAAGTAGCTACCTAGGATA 351
389 rSerLeuArgGlyGlnArgGlyThrLeuThrGlnAlaArgLysLysThr 406
|||||
352 TGCTTTAGAGATGAGACAGCGTGGCTCAGCAAAAACAGAAACAAAC 401
406 aValGlnIleProAlaProThrThrAlaLysGlnValArgGluPheLeu 422
|||||
402 TTATGCAATCCCAACCCCAACCACTGCTCCCGAGTAAGAGATTCTG 451
423 GlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAl 439
|||||
452 GGAACCGCGGGGTTTTCAGACTCTGATTCGCCGATTTCACACACTAGC 501

```

```

439 aalaproleuTyrrProleuThrLysGluLysGlyGlyPheSerTrpAla 456
|||||
502 AGCTCCCTTGTATCCACTACCAAGAGAGAGGGAATTCACCTGATAG 550
|||||
456 rogluHISgluLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAla 472
|||||
551 A.GAACATCAGCTAGACCTTTGAAACTCTCAAAAGGCACTCTCCAGGCT 600
|||||
473 ProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrrAla 489
|||||
601 CCGCATTTGGCCCTGGCCAGATTTAACAACCTTTCAACCCATATGATGA 650
|||||
489 pgluArglyGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProt 506
|||||
651 TGAAGAATAGAGTGGCAAGGAGGCGCTTACCAGGTTTGGGACCAT 700
|||||
506 TPATATGProValAlaTyrrLeuSerLysLysLeuAspProValAla 522
|||||
701 GGAAGCGCCGGTAGCTACTTAT..AAGAACTGACGCTGGCCAG 747
|||||
522 rGlyTrp 524
|||||
748 TGGATGG 754

```

seq_name: gb_AZ827524

seq_documentation_block: 700 bp DNA 20-FEB-2001
 LOCUS AZ827524
 DEFINITION 2M0104E02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 accession AZ827524
 keywords AZ827524.1 GI:12997432
 source house mouse.
 organism Mus musculus

REFERENCE
 AUTHORS Dunn, D., Moyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Keenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 unpublished (2000)

JOURNAL
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177

FEATURES
 source
 1..700
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0104E02"
 /clone_id="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (914732114[gb|AF129072.1]) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

alignment_block:
 US-09-171-553b-5 x AZ827524

alignment_scores:
 Quality: 960.50 Length: 233
 Ratio: 4.488 Gaps: 1
 Percent Similarity: 91.845 Percent Identity: 76.824

Align seg 1/1 to: AZ827524 from: 1 to: 700

```

416 LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpI 432
|||||
5 CGGACGATGCTGAGTCTTCTGGGACTGCTGCTTTGAGACTCTGAT 54
|||||
432 eProGlyPheAlaThrLeuAlaAlaProleuTyrrProleuThrLysGlu 449
|||||
55 TCCAGGCTTGGACACCTAGACGACCTCTATATCTTGGATAGAGAG 104
|||||
449 ysglyGlyPheSerTrpAlaProgluHISgluLysAlaPheAspAlaIle 465
|||||
105 GGGTTCCTTCGAGTGGAAAGAACACACAGAGAGCTTTGAGGCTATC 154
|||||
466 LysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValThr 482
|||||
155 AGTGTCTCTATATGACTGCCCCCGCTAGCATTACCAAGACTTGACTAA 204
|||||
482 sProPheThrLeuTyrrValAspGluArgLysGlyValAlaArgGlyVal 499
|||||
205 GCCTTTCGCTCTATATGAGAGAGAGAGCGGCTGATACCGAGGAGCT 254
|||||
499 euThrGlnThrLeuGlyProTrpArgArgProValAlaTyrrLeuSerLys 515
|||||
255 TGACACAGACACTGGACCTGGAGAGACCTGATGCTATTGTCAAAA 304
|||||
516 LysLeuAspProValAlaSerGlyTyrrProValCysLeuLysAlaIle 532
|||||
305 AAATTAGATCTCTGTTGCTAGTGGATGCCCCACATGCTGAAAGCTATTGC 354
|||||
532 aAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyLys 549
|||||
355 AGCATGACCTCTGTTGATCAAAAGATGCTGACAAATTGACATGGGACAC 404
|||||
549 snIleThrValIleAlaProHISAlaLeuGluAsnIleValArgGlnPro 565
|||||
405 AGGTGACTGTTGTGGCCCTCATCTAGAACTATCGTCCGACACCA 454
|||||
566 ProAspArgTrpMetThrAsnAlaArgMetThrHisTyrrGlnSerLeu 582
|||||
455 CCGTACAGATGATGACAAATGCGGATGACACACTATAGAGACCTGCT 504
|||||
582 uLeuThrGluArgValThrPheAlaProAlaAlaLeuAspProAla 599
|||||
505 GCTAAATGAGGCTGATACCTTTGGCCCGCTGCATCTCAACCCAGCTA 554
|||||
599 hrLeuLeuProGluGluThrAspGluProValThrHisAspCysHISglu 615
|||||

```



```

300 .....GlnLeuThrTrpThrArgLeuProG 308
753 CGAGAAAAGAGTTCAGAAAACATGAGCGTGCACCAAGAC..... 792
308 lnglYpHeLysAsnSerProThrIlePheAspGluAlaLeuHis...Arg 323
793 .....GAGACCAAGAGTCAGAAATCAGACGCTGCTGTGAGACCACGA 834
324 AspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTyr 340
835 GAGATCAGAAAATGAAATGTGGCCACCAAGTCCGGAATAATVtGAAAA 884
340 rValAspAspLeu.....LeuValIleGlnValThrL 351
885 CTTAGTNTTACCTGCGAACAAAAAAGCTCTTGCT.....C 925
351 ysglAspCysLeu...GluGlyThrLysAlaLeuLeuGlnLeuSer 366
926 AAGATCAATGTGTTAAGAGATACAAAAGGGGTCTAATACAG..... 969
367 AspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgI 383
970 .....AAGCTGAGTCTCTAGAAAGAGTCC 992
383 uValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrG 400
993 GGTGGCCTAACCTG.....TTGAAGC 1012
400 lValArgLysLysThrValValGlnIleProAlaProThrThrAlaLys 416
1013 ACCTGAAAAAGAGACTGTGTTCATCTCCACCGACGACGATGCCAAAA 1062
417 GlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIlePr 433
1063 CAGCTAAAAAGTCTCTGGGCACTCGGGCTTTTGCAGATTGTGATTCC 1112
433 oGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLysG 450
1113 AGTTTGCTTAAT.....TAAAGAGATTAAC 1138
450 lYglYPheSerTrpAlaProGlnHisGln.....LysAlaPheAspAla 464
1139 AGCCCTGTATAGAAAATATAAAACAACTTGATGGCCTTGATGCC 1188
465 IleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspValTh 481
1188 ATTAAGACTGCCCTGATGTATCCCACTATGGGACTCTAGATGTGAC 1238
481 lYsPheProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyV 498
1239 T.....GAGACAAAGCTATTGTCAAGAAAG 1264
498 alLeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSe 514
1265 TTCTTACCTAGAGATTGGACCTTGAAAAGACCTGTGGCATCTGTGT.. 1312
514 lYsLysLeuAspProValAlaSerGlyTrpProValLysLeuLysAlaI 531
1313 ..AAAAATTAGACCTGGTGGCTGTAGATGGCTGTGTGTGTGAC.... 1357
531 leAlaIleValAla..IleLeuValLysAspAlaAspLysLeuThrLeuG 547
1358 .....ATAGTGGCTTCTGTCAAGACGACGATTAATTGACCTGAG 1399
547 yGlnAsnIle.....ThrValIleAlaProHis..... 556
1400 ACAAAACTTGCCACATGTCTAAAAAGTGTGTTCCGCCCCACGACCGA 1449
557 .....AlaLeuGlnAsnIleValArg..GlnProProAspArgT 569
1450 TGGCTGACTAATGCTCTTGAAAACATTATCCACACGCTCCCTGACGAT 1499
569 rPmeThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuThrGlu 585
1500 GG.....ACACATTATCAGACCTTTTGTGTACTGAA 1531
586 ArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeuLeuP 602
1532 CGAGTGAACCTTCGCTCCCGCTGCTGCTCAAT.....CTCACTACTGCG 1575
602 roGlnGlnThrAspGluProValThrHisAspCysHisGlnLeuLeuIle 618
1576 CTGAGACTTCGCT.....ACTCATATTGTGTGACATTCCTGCA 1616
619 GluGlnThrGlyValArgLysAspLeuThrAspIleProLeuThrGly.. 634
1617 GAAGAACTGCTACTCCAAATGATCTAAAGATAGATCAAGCTTGCGCT 1666
635 GluValLeuThrTrpPheThrAspGlySerSerTyrValValGluGly 651
1667 GAGAGTTTGAGCTGTGACAGGATGCGATGACCTGAGGTTGAGGGTAA 1716
651 sArgMetAlaGly..Ala..AlaValValAspGlyThrArgThrIleTr 666
1717 GCGGAAGCGGGGACAGCAGTGCAGTGTGACAGAAAGAAAGATGATCTA 1766
666 rAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuMet 683
1767 GCCCAGCAGCTCCCTTAAGGAGCTTCAAGCCCTGAAGCTTAACCTTGG 1816
683 lAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyr 699
1817 CTTTAAATACAACT..... 1830
700 ThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAla..IleT 716
1831 .....CTAATATAT 1838
716 lYrGlnArgGlyLeuLeuThrSerAlaGlyArgGlnIleLysAsnLys 732
1839 ACAGACAAAGAGAGCTGTGACATCTGCA..... 1867
733 GluGlnIleLeuSerLeuLeuGlnAlaLeuHisLeuProLysArgLeuAl 749
1867 ..... 1867
749 alIleIleHisCysProGlnHisGln..LysAlaLysAspLeuIleSerArg 765
1868 .....GAAAAGACCTTAAGTGTGTGTGCTAA 1895
766 GlYAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAlaValAs 782
1896 AAAAATCAGATGGCAATCTAACCCGCCAGGACATCTTAAGAGCATGTAT 1945
782 nLeuLeuProIleIleGlnThrProLysAlaProGluProArgArgGlnT 799
1946 CCTG..... 1949
799 yTrpThrLeuGlnAspTrpGlnGlnIleLysLysIleAspIlePheSerG 815
1950 ..ACAGTCTAAAGACTATCAAGTTATAGACAAATTA..... 1983
815 uThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProH 832
1983 ..... 1983
832 lslYsglGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeu 848
1983 ..... 1983
849 GlYThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValle 865
1984 .....ACAGTGTAAAGAAAGAAACCTGTATGACATGAG 2015
865 uArgLeuProGlyValAlaAspSerValValLysHisCysValProCysG 882

```

```
2016 AGAAACTGAGAGGACACGAATATAGTCTCTCA..... 2049
882 InLeuValAsnAlaAsnProSerArgIleProGlyLysArgLeuArg 898
2050 .....TGA 2052
899 GLSerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPr 915
2053 GAAGCAGACCTGACATCTGAAAGATGACTTCACTGAAAGAAAAAT 2102
915 oAlaLysTrpGlyAsnLysTrpLeuLeuValPheValAspPheSerG 932
2103 GTGATG.....AATACCTTCTGG.....TTTTTGGAACGTTCTCA 2140
932 LTTTPValGluAlaTrpProThrLysLysGluThrSerThrValValAla 948
2141 GATGATTAAGCTTTCCTTATAAACTGAGACTACGATGATGTCATC 2190
949 LysLysIleLeuGluGluIlePheProArgPheGlyIleProLysVal 965
2191 AAGAAGATTATTAAAAAT...TTTCCAAGTTCGAGTGCCTCAAAAGCAAT 2237
965 eGlySerSpAsnGlyProAlaPheValAlaGluValSerGlnGlyLeu 982
2238 AGTGTCAATATAGTGTCTGCTTGTGTGCCCCAGTAAGTCAGGCTGTGG 2287
982 IaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTrpArgProG 998
2288 CCAAGTATTATAGAGTCAATACAAAGATTG.....TACAGACCTCA 2328
998 nSerSerGlyGluValGluArgMetAsnArgThrIleLysGluThrLeu 1015
2329 GAACTCAGACAGATAGAAATATATGATAAACTCT.AAACAGACCTTGA 2377
1015 hLysIleuThrThrGluThrGlyIleAsnAsp.TirMelAlaLeuLeuPr 1031
2378 CAAATTATATCTAGAGACTGGC...ACAGACTTACTTGTGGCCCTTTC 2424
1031 oPheValLeuPheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThr 1048
2425 CTTGCCCTATTAAACTGAGATACCTCCCTTGATAGGGTTTACTC 2474
1048 roTyrgIleLeuLeuTyrgIlyProProPheLeuAlaGluIleAlaPhe 1064
2475 TTTTAAAGATCTTTATGGGGCTCTATGCCCATCCTGTCTTAAATAT 2524
1065 AlaHisSerAlaAspValLeuLeu.SerGlnProLeuPheSerArgLeu. 1080
2525 GTGTTTAA...CCTATGTGTATATAATGATCTGATATCTAAGTTAT 2571
1081 LysAlaLeuGluTrpValArgGlnArgAla.....\TrpLysGln.L 1094
2572 AAGCCTTACAGGTGTCAGAAAGAGAGTGGTCACAACATGGCTACAGTG 2621
1094 euArgGluAlaTrpSerGlyLysPheGluValProHisArgPheGln 1110
2622 AACGAGCTGGGTACCCAGGACATCTTACCAGTTCCAGCCAGA..... 2665
1111 ValGlyAspSerValTyValArgArgHisArgAlaGlyAsnLeuGluTh 1127
2666 .....GATCGATCTGGCTACACCTGCTCATCTGAGACCCCTCAAGCC 2709
1127 rArgTrpLysGlyProTyrgLeuValLeuLeuThrThrPro 1140
2710 TCACAGAAAGGTCCTGACCTAGTTCTGTACTAATCCG 2749
```



```

53 GGGAGCTTTTCCAGCATCACTCAACGACCCGATGCCAGTTCTT 102
322 YSTPLCYSLYSEUSERSERGLYPROPTTYRGLUCLMETALALYS 338
103 GTTGGCTTGTCTATCTCCAGGCGCTCTATTATGAGGAATGGCTAGA 152
339 GLATGLYSPHEASNAVALTHIRYSGIUNHSARGASGLCYSTHTRPCL 355
153 GAAGGGAANTTCATGTGACCAAGACATAGAAATCAATGACATGAGGG 202
355 YSERATGASNLYSLEUTHRLTHRLGLUVALSERGLYSGLYTHRCYSI 372
203 GTTCCCAATAAGCTCACCCTCACTGAGTTTCCGGAGGGGACATGCA 252
372 IEGLYSVALPROPTROSERHISGLNHLSLEUCYSTYSETHRVALVAL 388
253 TGGAAAGACTCCCTCCATCCACCAACCTTTGCAATGACTATGTT 302
389 TYRGLUINALASERGLUNASGLINTYRLEUVALPROGLTYRASNAR 405
303 TATGAGCAGGCTTGAATAATCATGATTTAGTACTGGTTATGACTGAG 352
405 PTTPALACYASANTHRYGLYLEUTHRPROCYVALSERTHRSERVALPHEA 422
353 GTGGCATGCAATGACTGGGTAACTCCGTGTTCTTACACAGCTTTTA 402
422 SNGLSERLYSAPPHECYVALMETVALGNIILEVALPROATRYVALTYR 438
403 ACCAATCCAAAGATTCTGTGTGTCATGTCAGCTGCTCCCTGGGTACAC 452
439 TYRHSIPROGLUVALVALLEUASPSGLUTYRASPTRYRARGYRASNAR 455
453 TACCATCTCGAGGAGAGTAGTCATGATAGTATGATGCGCTACCCG 502
455 GPOLYSARGLUPROVALSERLEUTHRLVALVALMETLEUGLY 470
503 CTCAAAAGAGAGCGCCGTGACCTTAACCTTAGCGGTTATGCTAGA 549
seq_name: gb_gss:A2970955

```

```

seq_documentation_block:
LOCUS      A2970955      754 bp      DNA      GSS      27-APR-2001
DEFINITION 2M0244H16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
ACCESSION  A2970955
VERSION     A2970955.1 GI:13842182
KEYWORDS   GSS.
SOURCE      house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 754)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
TITLE       Unpublished (2000)
JOURNAL     Contact: Robert B. Weiss
COMMENT     University of Utah genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0244 row: H column: 16
            Seq primer: CATTGTAAACGACGCCACGT
            Class: plasmid ends
            High quality sequence stop: 754.
            Location/Qualifiers

```

```

source
1. 754
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0244H16"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-gold, M1-resistant, F-"
/notes="Vector: PMD42uv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (gl473211419blAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT  226 a 160 c 150 g 218 t
ORIGIN

```

```

alignment_scores:
Quality: 881.50      Length: 244
Ratio: 4.025        Gaps: 3
Percent Similarity: 89.754      Percent Identity: 67.623
alignment_block:
US-09-171-553b-6 x A2970955/rev ..

```

Align seg 1/1 to reverse of: A2970955 from: 1 to: 754

```

406 TPPLACYASANTHRYGLYLEUTHRPROCYVALSERTHRSERVALPHEAS 422
751 TGGGCTTGCAATAACCGCTCATTCCTGTGTGTCATGTCATGATTTTA 702
422 NGINSERLYSAPPHECYVALMETVALGNIILEVALPROATRYVALTYR 439
701 TGGTTCAAAGATTCTGTGTGTCATGTCATTCCTAGACCTCTGT 652
439 TYRHSIPROGLUVALVALLEUASPSGLUTYRASPTRYRARGYRASNAR 455
651 ATCATGATGATAGCTCCCTTTTAGACAAATTGACATCGGGTC...CGC 605
456 GPOLYSARGLUPROVALSERLEUTHRLVALVALMETLEUGLYLEUGI 472
604 TGGAGAGAGAGCCCATTAACCTTAAGTTCATGACATGATTAAGCATTAGG 555
472 YTHRALVALGLYVALGLYTHRLTHRALALALEUTHRLTHRGYPG 489
554 AGTACGCGGTGAGTAGTACGAGAACCGCTGCTTAATTAAGACCCGCC 505
489 LNGLNLEUGIULYSGLYLEUGIULNHSALALAMETCTHGLAS 505
504 AATACTAATGAA.....GAATCAGCTGACGATGATGATGTTGA 467
505 PLEUARGALALEULYSGLUSERVALSERASNLEUGIULNLSERLEUTHRS 522
466 TCTTGAACCTATAGAACAGTCTATACCAATTTAGAGAAATCTTAACCT 417
522 ERLEUSERGLUVALLEUGIUNASNRARGGLYLEUASPLEUAPHE 538
416 CCTGTCCGAAGTGCTGCTACAGATAGAGAGGATTAAGCTTATTTTC 367

```

539 LeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlyCysCysPheTy 555
 366 CTTAAAGAGAGGAGGAGCTGCTGCTCTAAAGAGATGTGTTTAA 317
 555 rValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 572
 316 TCTTACCATCTAGAGATATCAAGATTTATGCGCAACTGAGAGAAC 267
 572 ySLeuGluArgArgArgGluArgGluAlaAspGluGlyTyrPheGlu 588
 266 GCGTGAATACACGTAAAGGAGAAAGAGCAACAGAGATGTTCGAA 217
 589 GlyTyrPheAsnArgSerProTyrMetThrPheLeuSerAlaLeuTh 605
 216 ACCTGGTTAATAGTCCCTTGGCTCACACTCTCTCTCCACCATAGC 167
 605 rGlyProLeuValValLeuLeuLeuLeuThrValGlyProCysLeuI 622
 166 AGGACCTTGATCTACTCTTCTTCTTCTTGGCCCTGCATCC 117
 622 IaAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIle 638
 116 TTAATAGTTAGTACTTTTATTAAGAAAGGATTAATGACGTACAGGTT 67
 639 MetValLeuArgGlnGlnIleTyrGlnGlyLeu 648
 66 ATGCTACTAGGCAACATATCGGCTCTT 37
 seq_name: gb_BH059027

seq_documentation_block: 882 bp DNA 18-JUL-2001
 LOCUS BH059027
 DEFINITION RPI-24-326G3.TJ RPI-24 Mus musculus genomic clone RPI-24-326G3.
 ACCESSION BH059027 GI:14868646
 VERSION
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 882)
 REFERENCE
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akint, B., Levins, M.,
 Tesgaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 JOURNAL
 COMMENT Other_GSSs: RPI-24-326G3.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pje@jgemail.chi.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 326 row: G column: 3
 Seg primer: SP6
 Class: BAC ends.
 FEATURES
 source
 Location/Qualifiers
 1..882
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-326G3"
 /clone_1lb="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;

RPI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamH1 sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 248 a 183 c 201 g 250 t
 ORIGIN

alignment_scores:
 Quality: 865.50 Length: 309
 Ratio: 3.652 Gaps: 6
 Percent Similarity: 76.699 Percent Identity: 58.252

alignment_block:
 US-09-171-553b-6 x BH059027 ..

Align seg 1/1 to: BH059027 from: 1 to: 882

297 GlyLnaArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIleAsnSe 313
 2 GGGCAAGAGAGCTGTCACACGATTAAGAGAGAGCTTTATGCGCTTACAG 51
 313 rThAspProAspAlaThrSerSerCysTyrPheCysLeuSerSerGlyP 330
 52 CTCGATCCAGATGTCACATCATCTGCTGTTATTCCTATCATCAGGCC 101
 330 rOPrOTyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLys 346
 102 CTCCTACTCTGAGAGGGTTCCTTTATTAAGATTTTAAATTAACAGT 151
 347 GluHisArgAsnGlnCysThrTyrPglySerArgAsnLysLeuThrLeu 362
 152 AGTCATGCC...TTATGCTCATGGGCAACAGAAAGAAATTAACCTGTA 198
 363 ...ThrGluValSerGlyLysGlyThrCysIleGlyLysAlaProPro 378
 199 CTGAAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
 378 eRhISGlnHisLeuCysTyrSerThrValValTyrGlnGlnAlaSerGI 394
 249 CCCACCAACACCTATGTGTGGAGCTTATATCTGTATCCAGAGCAATAC 298
 394 uAsnGlnTyrLeuValProGlyTyrAsnArgTyrTyrPAlaCysAsnThrG 411
 299 CAACATATTACCTTGGCCATATCCAGTTGGTGGGCGCTGTAACAG 348
 411 lYleuThrProCysValSerThrSerValPheAsnGlnSerLysAspHe 427
 349 GGCCTCACCCCATGTGTATCACTAGTGTGCTTGTGACCCCTGTGATTTT 398
 428 CysValMetValGlnIleValProArgValTyrTyrHisProGluLys 444
 399 TGTATCATGCTTCATCTCTATCTGCATGATATTAATTCCTGCCTCCG 448
 444 lValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgLysP 461
 449 CCGTGAAGATGATATATACAGTAAG...AAGTTGAAGAGATCCAG 492
 461 aSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAlaValGlyAl 477
 493 TCTTCATTAATCTAGTATGTTAAAGGGGGGCGCTTACAGA.GGAATA 541
 478 GlyThrGlyTyrAlaAlaLeuThrGlyProGlnGlnLeuGluLysGI 494
 542 GGAATGGGTTTAGCAGCCTTGCTAGAGGTAGACAGAAATTCACACTT 591
 494 yLeuGlyLeuLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysG 511
 592 CTTAAGAGAA.....GCTATAAATGAAGATTTAAGCATGTCACAAA 632
 511 lSerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValVal 527
 633 AATCTGTGATGATTAAGAAATCTTGCAGCTCCCTGTCAGAAAGTAGTA 682

528 LeuClInaSnArGArGgLyLeuAspLeuLeuPheLeuArgLysGlyLeu 544
 683 TTACAAAACAGAGAGGTTAGATTATGTTTAAAGGAGAGGTGACT 732
 544 uCysAlaLeuLeuGlyCysCysPheTyrValAspHisSerGlyA 561
 733 ATGTGACGCTTAAAGAGAAATGCTGCTTCTATGACAGCACACGGGAA 782
 561 laileArgAspSerMetAsnLysLeuArgLysLysLeuArgArgArg 577
 783 TAGTCAGAGATCTATGACAAATGAGGAAAGACTGGAAAGGGA... 829
 578 ArgGluArgGluAlaAspGlnGlyTrrPheGluGlyTrrPheAsnArg 594
 830TGGTTCAGAGTGTGTGACTTAAATC 855
 594 rProTrrMetThrThrLeuLeuSer 602
 856 TCCATGATGACACACTTACTTCT 880

seq_name: gb_gss:BH122287

seq_documentation_block:

LOCUS BH122287 774 bp DNA GSS 19-JUL-2001
 DEFINITION RPI-24-288P13.TJ RPI-24 Mus musculus genomic clone RPI-24-288P13
 , DNA sequence.

ACCESSION BH122287
 VERSION BH122287.1 GI:14965799

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 774)

AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinet, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregiorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPI-24

JOURNAL Unpublished (1999)

COMMENT Other-GSSs: RPI-24-288P13.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
 page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
 Plate: 288 row: P column: 13
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers

1..774

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPI-24-288P13"

/clone_lib="RPI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBACK; site_1: BamHI; site_2: BamHI;
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBACK cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT

221 a 162 c 172 g 219 t

alignment_scores:
 Quality: 849.50 Length: 262
 Ratio: 3.776 Gaps: 2
 Percent Similarity: 85.878 Percent Identity: 62.214

alignment_block:
 US-09-171-553b-6 x BH122287

Align seg 1/1 to: BH122287 from: 1 to: 774

354 TrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyTh 370
 2 TTTGGGAGAAACAAAGTTGACTCTGCGACAGCTTTCAGAGAGAGGCT 51
 370 rCysLleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThr 387
 52 TTGTTGGGCGAGGACTCTGAGATAAAGGCGACCTGTATATCAGACC 101
 387 alValTyrGluGlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsn 403
 102 AGAATCATCAGTCTAGCAAAAGTGTGATGATGATGATGATGATGATG 151
 404 ArgTrrPrrAlaCysAsnThrGlyLeuThrProCysValSerThrSerVa 420
 152 ACAAGTGGGCTGCTGCAATACCGGTCTGCTGCTGCTGCTGCTGCTG 201
 420 lPheAsnGlnSerLysAspPheCysValMetValGlnIleValProArgV 437
 202 TTTTAATAGTTCCAAGATTTCTGCTATTTGGTTCACCTTATTCCTAGAC 251
 437 aLTYrTYrHisProGluGluValValLeuAspGlnTyrAspTyrArgTyr 453
 252 TCTGATATCATGATGATAGAGCCCTTTTATGACAAATTTGAGCATGGGTC 301
 454 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 470
 302 ...CGCTGAGAGAGAGAGCCGTTACCTTACTTGGCAGCTTCATTAGG 348
 470 yLeuGlnThrAlaValAlaGlyValGlyThrGlyThrAlaAlaIleThrG 487
 349 ATTGAGAGTACGGCTGAGTACGAGTACGAGACCGCTGCTTATTTAGCA 398
 487 lYrProGlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaIleMetThr 503
 399 CCCCCAATAGTATGAA.....CACTACGTCGACGATGATGAT 436
 504 GluAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnLysSer 520
 437 GTTGTATCTTGAACCTATGACAGTCTATTAACCAATTTAGAGAAATCTT 486
 520 uThrSerLeuSerGluValAlaLeuGlnAsnArgArgGlyLeuAspLeu 537
 487 AACTTCCTGCTCGAAGTGTGCTACCAATTAAGAGGATTTAGACTTAT 536
 537 eupHeLeuArgGluGlyLeuCysAlaAlaLeuLysGlnLysCys 553
 537 TATTCCTTAAAGACGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGA 586
 554 PheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 570
 587 TTTTATGTTGACATTCATGATGATTAATCAAGATTTATGAGCCCAACTT 636
 570 gLysLysLeuGluArgArgArgGluArgGluArgGluAlaAspGlnGlyTrrP 587
 637 AGAAGCGCTGATATACGTGAAGAGAAATAGAAACCAAGAGATGAT 686
 587 heGluGlyTrrPheAsnArgSerProTrrMetThrThrLeuLeuSerAla 603
 687 TCGAAGCTGGGTTAATAGTCCCTTGGCTGACACCACTCTCTCCACAC 736
 604 LeuThrGlyProLeuValAlaLeuLeuLeuLeu 615

seq_name: gb_gss:A2850764

seq_documentation block:

LOCUS A2850764 641 bp DNA GSS 21-FEB-2001

DEFINITION ZM0152C18R Mouse 10kb plasmid U06C1M library Mus musculus genomic

ACCESSION A2850764

VERSION A2850764.1 GI:13036088

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 641)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

Unpublished (2000)

JOURNAL COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA

Tel: 801 585 5506

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0152 row: C column: 18
Seq primer: CACACAGCAACAGCTATGACCC
Class: plasmid ends
High quality sequence stop: 641.

BASE COUNT	197 a	134 c	123 g	187 t
ORIGIN				

alignment_scores:		
quality:	776.50	218
ratio:	3.982	2
Percent Similarity:	89.450	Percent Identity: 66.514

```
alignment_block:
```

US-09-171-553B-6 X AZ850764/rev

Align seg 1/1 to reverse of: A2850764 from: 1 to: 641

```

416 ValSerhhSerValPhehnsIndSerIysaspPhecyValMetValIgi 432
639 GTGCTCATGTCTGTTTTTAATAGTTCACAAAGATTTCGTCAATTTGGTTCA 500
449 nIleValProArgValTyTyrHisProGluGluValValLeuAspLut 449
589 GCTATTCCAGACCTCTGTATCATGATGATAGTCACCTTTTATGACAAAT 540
449 yrasPtyIarGlyIrasnarGProLysArgGluProValSerLeuThleu 465
539 TTGAGCATCGGGTC...CGCTGGAAGAAAGAACCCGTACCTTACCTTGG 493
466 AlValMetLeuGlyLeuGlyIthrAlaValGlyValGlyThrAl 482
492 GCGATCTTATTAAGATTGGAGTAGGAGCGGTGAGTAGGTACAGCAACCC 443
482 alaLeuIleThrGlyProGlnGluLeuGluIubSerGlyLeuIubH 499
442 TGCTTATATTAAACCCCAATACATGAA.....CAACTAC 405
499 isAlaIaIaMetThrGluAspLeuAlaGAlaLeuIubGluSerValSerAsn 515
404 GTGCAGCTAAGCATGTGTGATTAGAACTATGAAACAACATCTATTAACCAA 355
516 LeuGluIubSerLeuThrSerLeuSerGluValAlaLeuGlnAsnArgar 532
354 TTGAAGAAATCTTAACTTCCCTGCTCCGAAGGTGCTACAGAAATAGAG 307

```

seq_name: gb_gss:AZ813042

```
seq_documentation_block:
```

LOCUS	Accession	Size	Library	Genomic
DEFINITION	AC013042	0.23 kb	mus	musculus
	2M0080E11	Mouse	plasmid	U061M
	clone U06C2M0080E11	F, DNA sequence.		

ACCESSION	AZ813042	
VERSION	AZ813042.1	GI:12982881
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE
Dunn, D., Aoyagi, A., Barber, M., Beacom, T., Duval, B., Hamil, C.,
1 (bases 1 to 625)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
Department of Health Services
University of Washington

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Seq primer: CGTTGTAAMCGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 625.

FEATURES
SOURCE

532 gggluLeuaspLeuLeuPheLeuArgIGluGluValLeuLeu 549
304 GGGATTAGACTTATTATTCCTTAAAGAAGAGAGACTCGTCTCCCTAA 255
549 ysgIuGluLucysCysPheTyrValAspHisSerGluAlaLeuArgAspSer 555
254 AAGAGAAAGCTGTTTGTGTGTGACCATTCAGGAGTAAATCAAGATTC 205
566 MetAsnLysLeuArgLysLysLeuLuarGArgArgArgGluArgGluAl 582
204 ATGGCCAAACTTAGAAGAACCTCAATATCTTAAAGAAGAAGAAG 155
582 aaSpGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyrPheThr 599
154 CCACAAAGGTGGTTTGAAAGGTGGTTTAAATAAAGTTCCTTGCGCTACCA 105
599 hrLeuLeuSerAlaLeuPheArgGlyProLeuValValLeuLeuLeuLeu 615
104 CTCTCCTCTCCACCAATAGAGAGACTTAAATACACTATGCTTTTGCTT 55
616 ThrValGlyProCysLeuLeuAsnArgPheValAlaPheValArgGluArg 63
54 ACTTTGGGCCCGCATCTTATATAGGTGGAGCTTTTATTAGGAAG 5

632 gVal 633

4 GATA 1

name: gb_gs

documentati

[illegible]

314 raspProaspalathr

```

54  TGAATCCAAAGCGCTCTCTAGGAGACTGCTGGCATATGCGTGTCTCTGGGGTCCG 103
331  TGTATTGTGTGUGUyMetalALySGUArGySPheasNalThrLYSGU 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104  CTTATTATTGAAGAATCGCTTCATATGGAATTTTCAACAGAACAGAGC 1533
348  HIsArgasnGInCySThrTrpGlySerArgAsnLysIleuThrIleuArg 364
|||  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154  CAT...ACTTCCCTCTCTGTGGGTAACGGACAAAAACCTGCACTGATGA 2000
364  uAlSerGlyLys.....GlyThrCysIleGlyLysAlaProPserH 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  AGTATCCGGCAGGAATCCAGGTCTCTGTATAGGTAACCCCACTTCACATC 250
379  IsGInHisLeuCySTyrSerThrValValTyrgInGInAlaSerGluAsn 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251  ACAAAACACTATGCGGACAAATTCAGTCCGTGTCACAGAACGGAAGCTGAT 3000
396  GInTyrlleuValProGlyTyrrAsnArgTrpTrpAlaCysAsnThrGly 412
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301  TACTATCTTGACCTCCCGGTGATGGTGGGCTTGCAATACAGACT 3500
412  uThrProCysAlaSerThrSerValPheAsnGInSerLysAspPheCysV 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351  TACTCCCTGTGTATCAACTAGGTTTAAATATCATCTCATGATTTTCGTG 4000
429  alMetValGInIleValProArgValTyrrTyrrHisProGInGluVal 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401  TCATGATCCACAGCTTACCCCGGTATATATATATACCCGTGATCCAGTTTA 4500
446  LeuAspGluTyrrAspTyrrArgTyrrAsnArgProLysArgGluProValSe 4620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451  GAACACAGACTATGGGGCCGG.....CGGTCAAAAAGAACCAACTTAC 4940
462  rLeuThrIleuAlaValMetIleuGlyLeuGlyThrAlaValGlyValGly 4790
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495  TTTAACTCCGTGCTGATTCATGAGGAATAGGTAATGGCAGTGGAGTGGGA 5440
479  hrGlyTyrrAlaIleValIleThrGlyProGInGInLeuGluLysGlyL 4955
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545  CGGGAGTGTCAAGCTTTCGATGAGAACGGAAAGACAG.....GGAA 5820
495  euGlyGluLeuHisAlaAlaMetThrGluAspLeuAlaGlyAlaLeuLysGlu 5110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583  TTCACTCTTTAGGGATGCTGTCAATGAAAGCACTTACGGGCAATGAGAGAG 6320
512  SerValSerAsnLeuGInGluSerLeuThrSerLeuSerGluValIle 5280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633  TCCATTTGACGCTTATGAAAAATCTTTTGAACCTCCCTGTCTGACGTAGTTT 6820
528  uGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuC 5450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683  ACAGAAACGAGAGAGGTGTGATGTGCTGCTTAAGACAGAGAGAGAGCTGT 7320
545  ysaIaAlaLeuLysGluGlyCysCysPheTyrrValAspHisSerGly.A 5610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733  GTGCTGCCCTTAAGAAAGACAGTGCCTTCAATTCACATCATACAGGAAT 7820
561  laIleArgAspSerMetAsnLysLeuArgLysLysLeuArgGlyGlyG 5760
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783  TAGTTAGAGACTCTCATATGACAACTGGACAGAAAGATTTTCACCGAAGG 8330
577  ArgArgGluArgGluAlaAspGInGlyTyrrPheGluGlyTyrrPheAsnA 5930
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
833  AACACGGGAAACCGATGCCCAACGGGGTACGTGAGTGGGTTGAATCCA 8820
593  rg 593
||
883  GA 884

```

LOCUS	AZ356870	644 bp	DNA	GSS	02-OCT-2000
DEFINITION	M0099FF09F Mouse 10kb plasmid UGCGIM library Mus musculus genomic clone UGCGIM0099F09 F, DNA sequence.				
ACCESSION	AZ356870				
VERSION	AZ356870.1	GI:10470556			
SOURCE	GSS.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 644)				
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid insets				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 302, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0098 row: F column: 09 Seq primer: CGTTGTAACGACGACGCCAGT Class: plasmid end High quality sequence stop: 644. Location/Qualifiers 1..644				
FEATURES					
source	1..644				

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UMGC1M0098P09"
/clone_lib="Mouse 10kb plasmid UMGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/vector="PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114[gb|AF129072.1]), a copy-number inducible derivative of Plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      185 a      136 c      133 g      190 t
ORIGIN

alignment_scores:
    quality: 740.50      Length: 215
    Ratio: 3.897      Gaps: 2
Percent Similarity: 88.372      Percent Identity: 66.977

alignment_block:
US-09-171-553B-6 x AZ356870/rev ..

Align seg 1/1 to reverse of: AZ356870 from: 1 to: 644

```


seq_name:	gb_gss:AZ376517
seq_documentation:	block:
LOCUS	AZ376517 610 bp DNA
DEFINITION	1M0130J20F Mouse 10kb plasmid UGCG1M library Mus musculus genomic clone UGCG1M0130J20 F, DNA sequence.
ACCESSION	AZ376517
VERSION	AZ376517.1
KEYWORDS	GI:10490217
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scleroognathi; Muridae; Murinae; Mus 1 (bases 1 to 610)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)

Contact: Robert B. Weiss
University of Utah
Genome Center
Rm. 308, Biomedical
Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 385 5606
Fax: 801 585 7177

Email: ddunne@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0130 row: J column: 20
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 610.

1. .610

```
/db_xref="taxon:10090"  
/clone="UUC1M0130J20"
```

```
/clone_11b="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
```

```
/lab_host="E. coli strain XL10-Gold, T1-resistant"
/note="Vector: PWD42nv; Purified genomic DNA f
```

musculus C57BL/6J (male) was obtained from the Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares>) was hydrodynamically sheared by repeated passage

0.005 inch orifice at constant velocity. The sample was blunt-end-repaired with T4 DNA polymerase

polynucleotide kinase. Adaptor oligonucleotide ligated to the blunt ends in high molar excess

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was removed by

of pWD42 (gi14732114|gb|AF129072.1), a copy-number of 1000.

inducible derivative of plasmid RI. The vector with adaptors complementary to the insert adap

purified, adapted mouse DNA was adapted vector DNA, and transformed into

chemically competent *E. coli* XL10-Gold (Stratagene) and selected for ampicillin resistance."

7 20T 6 02T 3 15T

730.50	Length:	204
3 070		

3.570 Gaps: 2
90.196 Percent Identity: 67.157

2376517/row

Reverse of: A2376517 from: 1 to: 610

```

reverse of. AZ3/031/ 110m: 1 CO: 010
BrCArGVa]TvrTvrH;sBrGc\wclwv}w}fowb 447

```

[illegible]

ArctTyrAsnArgProIysArgC|uBroYs}SerIauT 464
ccnnnncccccnnnnnnnnnnnnnnnnnnnnnnnnnn 559

[illegible]

ATTCG|VTHrA|AYa|C|vYa|C|wThrc|w 480
ACCGGC...CCCTGGAGAGAGGCCCGIACCIIAM 512

ATTAGGATTAGGAGTACGGCTGCACTACCTACCGCA 463

[illegible]

TTAAGACCCCAATACTATGAA GA 434

```
MethBrG]ASn[eAUrA]a[eUvSG]"SeryaJS 514
```

[illegible]

423 ACTAGCTGCAGCTATGATGTTGANTCTAGACTATAGACAGCTCTATAA 374
 514 eTaSnLeuGlUGluSerLeuThrSerLeuSerGluValValLeuGlnAsn 530
 373 CCAAAATTAGAAAGATCTTTAACTTCCCTGCGAAGTGGTCTACAGAAAT 324
 531 ATGATGlyLeuAspLeuPheLeuArgGluGlyGlyLeuCysAlaAl 547
 323 AGAAGGGATTTAGACTTATTATCTTAAGAAGAGGAGCTGTGTCTGC 274
 547 AleuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIleArgA 564
 273 CCAAAAGAAGATGTTGTTTATGTTGACCATTCAGAGTAATCAAGAAG 224
 564 sPSeuLysLysLeuArgLysLysLeuGluArgArgArgGluArg 580
 223 ATCTATGGCCAAACTTAGAAGACGCTAGATATAGCTAAAGAGAAAGA 174
 581 GluAlaAspGluGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMe 597
 173 GAAGCCCAACAGATGTTCCAAAGCTGTTAAATATATATATATATATAT 124
 597 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeu 614
 123 CACCACCTCTCTCTCCACCATAGACAGCTTGTATACACTATATCTTT 74
 614 euleuthrValGlyProCysLeuIleAsnArgPheValAlaPheValArg 630
 73 TGCTTACTTTGGCCCTGCATCCTTAATTAATAGTTAGCTTTTATTTAGA 24
 631 GluArgValSer 634
 23 GAAGAGCTAAT 12
 seq_name: gb_gss:AZ430249
 seq_documentation_block:
 LOCUS AZ430249 591 bp DNA GSS 03-OCT-2000
 DEFINITION 1M0214G11R Mouse 10kb plasmid U08C1M library Mus musculus genomic
 clone U08C1M0214G11 R, DNA sequence.
 ACCESSION AZ430249
 VERSION AZ430249.1 GI:10554262
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 591)
 REFERENCE
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0214 row: G column: 11
 Seq primer: CACACGAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 591.
 FEATURES
 1. 591 Location/Qualifiers
 source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"

/clone="U08C1M0214G11"
 /clone_lib="mouse 10kb plasmid U08C1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-gold, T1-resistant, F-"
 /note="Vector: pMD22nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g1147321149b/AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 165 a 125 c 115 g 186 t
 ORIGIN
 alignment_scores:
 Quality: 729.00 Length: 194
 Ratio: 4.096 Gaps: 1
 Percent Similarity: 91.753 Percent Identity: 71.134

alignment_block:

US-09-171-553B-6 x AZ430249/rev ..

Align seg 1/1 to reverse of: AZ430249 from: 1 to: 591

455 ArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe 471
 588 CCCTGGAAAGAGAACCCGTTACCTTATGAGTTTATTAATAGATTT 539
 471 uGlyThrAlaValGlyAlaGlyThrGlyThrAlaAlaLeuIleThrGlyP 488
 538 GGGAGTAGACAGCTGGAGTAGACAGAACCCGCTTAAATTAACACCC 489
 488 tGcGlnGlnLeuGluLysGlyLeuGlyLeuHisAlaAlaMetThrGlu 504
 488 CCCAATTAATGAA.....GAGCTACGTCGAGCTATGGATGTT 451
 505 AspLeuArgAlaLeuLysGluSerValSerAsnLeuGluGluSerLeuTh 521
 450 GATCTTAGACCATAGAACAGCTATTAACCAATTAGAAGATCTTTAAC 401
 521 rSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuP 538
 400 TTCCCTATCCGAAAGTGCTGCAAAATAAGAGGGGGTGTAGCTTATATAT 351
 538 heLeuArgGluGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGly 554
 350 TCCTTAAGAGAGGAGACTGTGCTCCCTTAAGAAAGAAATGTTGTTT 301
 555 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArg 571
 300 TATGTTGACCATTCAGAGATATCAAGATCTTATGCGTAAACTTAAGA 251
 571 sLysLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheG 588
 250 AGCGCTAGATATACGTAAAGAGAAAGAGAAAGAGAGAGAGAGAGAGTTG 201
 588 uGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeu 604
 200 AAAGCTGTTTATTAATAGTCCCTTGGCTACACACTCTCTCTCCACTATA 151

```

605 thrGlyProLeuValValLeuLeuLeuLeuTrpAlaGlyProCysLeu 621
    ::::::::::::::::::::|||:::
150 GCAGAGACCTTAAATTACACTTATGCTTTTCTTACTTTTGGCCCATGCAT 101
    ::::::::::::::::::::|||:::
621 uIleasnArgPheValAlaPheValArgGluArgValSerAlaValGlnI 638
    ::::::::::::::::::::|||:::
100 CCTTATATAGTAGTAGCTTTTATATAGAGAAAGATATAAACGACATCCAG 51
    ::::::::::::::::::::|||:::
638 lMetValLeuArgGlnGlnTyrGlnGlyLeu 648
    ::::::::::::::::::::|||
50 TTATGTACTAAGCACAATATCGGTCCTT 19
    ::::::::::::::::::::|||
seq_name: gb_gss:A2337993

```

seq_documentation_block:	
LOCUS	AZ337993 673 bp DNA 29-SEP-2006
DEFINITION	1M0608F2R Mouse 10kb plasmid UUC1M library Mus musculus genomic
ACCESSION	clone UUC1M0068F2 R, DNA sequence.
VERSION	AZ337993
KEYWORDS	AZ337993.1 GI:10410826
SOURCE	GSS.
ORGANISM	house mouse.
	Mus musculus

REFERENCE
AUTHORS
Dunn, D., Koyagi, A., Barber, M., Beacom, T., Duval, B., Hamill, C.,
Eukariyota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 673)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.

plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG,
84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0068 Row: F Column: 24
 Seq primer: CACACAGGAACACGCGTATACC
 Class: plasmid ends
 High quality sequence stop: 673.

FEATURES	Location/Qualifiers
source	1. .673

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U08C1M0068P24"
 /clone_lib="Mouse 10Kb plasmid U08C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv: Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivatized
 of PWD42 (q114732114[9b]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells.

BASE COUNT	and selected for ampicillin resistance."
197 a	120 c
167 g	188 t
1 others	

alignment_scores:	
Quality:	727.50
Ratio:	3.638
Percent Similarity:	87.719
	Percent Identity: 59.211

```
alignment_block:
US-09-171-553B-6  x  A2337993  .
```

Align seg 1/1 to: AZ337993 from: 1 to: 673

```

420 ValPheAsnGlnSerLysAspPheCysValMetValGlnIleValProL 43
    |||||..... |||||..... |||||..... |||||..... |||||
1 GTTTATTATCATCTCATGATTTTGTGTGCATGATCCAGCTTACCCG 50
436 gValIyrrTyrrHisProGluGluValValLeuAspLuTyrrAspTyrrArgT 45
    |||||..... |||||..... |||||..... |||||..... |||||
51 CGTATTATTATCACCTCGATCCAGTTTA.... GAAGAAACCTATGCGT 94
453 yrrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeu 46
    :|||: |||||..... |||||..... |||||..... |||||.....
95 GCCGCGCGTCAAAAAGACAAACCAATTTACTTTRACCTTGGCTGCATTCA 14
470 GlyLeuGlyThrAlaValAlaGlyValGlyThrGlyThrAlaAlaLeuIle 48
    |||||..... |||||..... |||||..... |||||..... |||||
145 GGAATAGTAGTATAGCGAGTGGAGATGGGACGGAGCTCAGCTTGATAGA 19
486 rGlyProGlnGlnLeuGlyLysGlyLeuGlyGlnLeuHisAlaAlaMet 50
    ||| ||| |||||..... GGATTTCAGCTTTGAGGAGCTTGCTGCA 23
195 AGGAAGACAG..... GGATTTCAGCTTTGAGGAGCTTGCTGCA 23
503 hrGluAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnGluSer 51
    :|||: |||||..... |||||..... |||||..... |||||.....
233 ATGAAGACCTGACACCAATATAGAGAGTCCATTGACGCTTTAGAAAATCT 28
520 LeuThrSerLeuSerGlnValValLeuGlnLysAsnArgGlyLeuAspLe 53
    |||||..... |||||..... |||||..... |||||..... |||||
283 TTGACCTCCCTGTCTGAGGTAGTATTACAGAAACAGAGAGGCTTGATTT 33
536 uLeuPheLeuArgGlnGlyGlyLeuCysAlaAlaLeuLysGlnGluCysC 55
    |||||..... |||||..... |||||..... |||||..... |||||
333 GTTGCTTCCTAAGGAGAGGACGTGTGTCGCCCTTAAGAAAGTGCT 38
553 ysrPheTyrrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 56
    |||||..... |||||..... |||||..... |||||..... |||||
383 GCTTCTATGACGATCATACAGGAATATGTTAAGACTCTATACAGAACTG 43
570 ArgLysLysLeuGlnArgArgArgArgGluArgGluAlaAspGlnGlyTr 58
    |||||..... |||||..... |||||..... |||||..... |||||
433 AGAGAAAGATTAGAGCAAAAGGAACGGGACGGGATGCTCAACGGGGTG 48
586 pPheGlnGlyTrpPheAsnArgSerPro... TrpMetThrThrLeuLeus 60
    |||||..... |||||..... |||||..... |||||..... |||||
483 GTTTGAGTCGCGGTTTGATACACAAACCTCTTGATTAACCTCTTAATTT 53
602 eAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGly 61
    |||||..... |||||..... |||||..... |||||..... |||||
533 CCGGTGATAGCGGACCAATCTCTTATGATATGCTTAGCTNATGTTTGGC 58
619 ProCysLeuIleAsnArgPheValAlaPheValAlaArgGluArgValSerAl 63
    CCTGTGTATATATATAGAGAAATGCTTTCACAGAGTAAATATGATAC 63
583 CTTGTGTATATATATAGAGAAATGCTTTCACAGAGTAAATATGATAC 63
635 aValGlnIleMetAlaLeuArgGlnGlnIntyrgln 64
    |||||..... |||||..... |||||..... |||||..... |||||
633 AGTAAACCTCATGGTTCTTCAAAAGCAATATCA 66
seq_name: gb_gss:A5642336

```

seq. documentation block:	624 bp	DNA	GSS	14-DEC-2000
LOCUS	A2642336			
DEFINITION	IM0505M18F Mouse 10kb plasmid U06C1M library	Mus musculus genomic		
ACCESSION	clone U06C1M0505M18 F, DNA sequence.			
VERSION	A2642336			
KEYWORDS	A2642336.1	GI:11768843		
SOURCE	GSS.			
ORGANISM	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 624)			
	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,			
	Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly			
	,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.			
	and Wright,D., Weiss,R.			
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb			
	plasmid inserts			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Robert B. Weiss			
	University of Utah Genome Center			
	University of Utah			
	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT			
	84112, USA			
	Tel.: 801 585 5606			
	Fax: 801 585 7177			
	Email: ddunn@genetics.utah.edu			
	Insert Length: 10000	Std Error: 0.00		
	Plate: 0505	row: M	column: 18	
	Seq primer: CGTGTAAACGACGCGCACAT			
	Class: plasmid ends			
	High quality sequence stop: 624.			
FEATURES	Location/Qualifiers			

```

SOURCE
1. .624
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U061M0505M18"
/clone_id="Mouse 10kb plasmid U061M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42n; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMW42 (g1147321141gD/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

[illegible]

alignment_block:

Align seg 1/1 to: AZ642336 from: 1 to: 624

[illegible]

seq_documentation_block:	659 bp	DNA	GSS	25-JAN-2001
LOCUS	AZ741412			
DEFINITION	RPCT-24-74C13.TV	RPCT-24	Mus musculus genomic clone	RPCT-24-74C13
ACCESSION	AZ741412			
VERSION	AZ741412.1	GI:12518156		
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 659)			
AUTHORS	Zhao, S., Nierman, M., Malek, J., Shatsman, S., Akınrit, B., Levins, M., Tsegaye, G., Geer, K., Kroll, M., Shwartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P., and Fraser, C. M.			
TITLE	Mouse BAC End Sequences from Library			
JOURNAL	Unpublished (1999)			
COMMENT	Other-GSSs: RPCT-24-74C13.TV			

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 74 row: C column: 13
 Seg primer: SP6
 Class: BAC ends.

FEATURES

Source Location/Qualifiers

1.659
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-74C13"
 /clone_id="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 186 a 132 c 152 g 189 t
 ORIGIN

alignment_scores:

Quality: 719.50 Length: 220
 Ratio: 3.868 Gaps: 2
 Percent Similarity: 84.545 Percent Identity: 63.182

alignment_block:

us-09-171-553b-6 x AZ741412

Align seg 1/1 to: AZ741412 from: 1 to: 659

```

355 GlySerArgAsnIlyLeuThrLeuThrgluValSerGlyIlyThrCy 371
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
10 GGGAGAAACAGAAAGTTGACTCTAGCAGCACTTTCAGAGAGGCGCTTG 59
371 sllEgIlylAslAproProSerHisGlnHisLeuCySlyrSerThrVal 388
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
60 TCTGGCGCGGTACTCAGTTAAAGGCGCACCTCTGTAATCAGACCCAGA 109
388 aLYrGlulAlaSerGlnAsnGlnIlyrLeuValProGlyTYrAsnArg 404
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
110 ACATCCAGCTAGCAAAAGTGTGCAATCTAGTGTGCCCCCTTAGACACA 159
405 TrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerValph 421
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
160 GTGTGGCTTGCATACCGGTCTCACCCTGTGTGTCTATGCTGTTT 209
421 eAsnGlnSerIyAspPheCysValMetValGlnIlyValProArgValT 438
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
210 TAAATAGTCCAAAGATTCTGATTTGGTTCACCTATTCTTAGACTCC 259
438 YrTYrHisProGluValValLeuAspGluTYrAspTYrArgTYrAsn 454
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
260 TGTATCATGATGATAGCTCCTTTTAAACAAATTGAGACATCGGGTC... 306
455 ArgProIyAsrGluProValSerLeuThrLeuAlaValMetLeuGlyLe 471
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
307 CGCTGGAGAAAGAACCGCTTACTTAACTTGGCAGTCTCATTTAGGATT 356
471 uGlyThrAlaValAlaGlyThrGlyThrAlaAlaLeuIleThrGlyP 488
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
357 GGGAGTAGCGGTGAGTAGTACAGAACCATGCTTATTAATTAAGACC 406
488 rGlnGlnLeuGluIlyGlyLeuGluIlyLeuHisAlaIleMetThrGlu 504
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

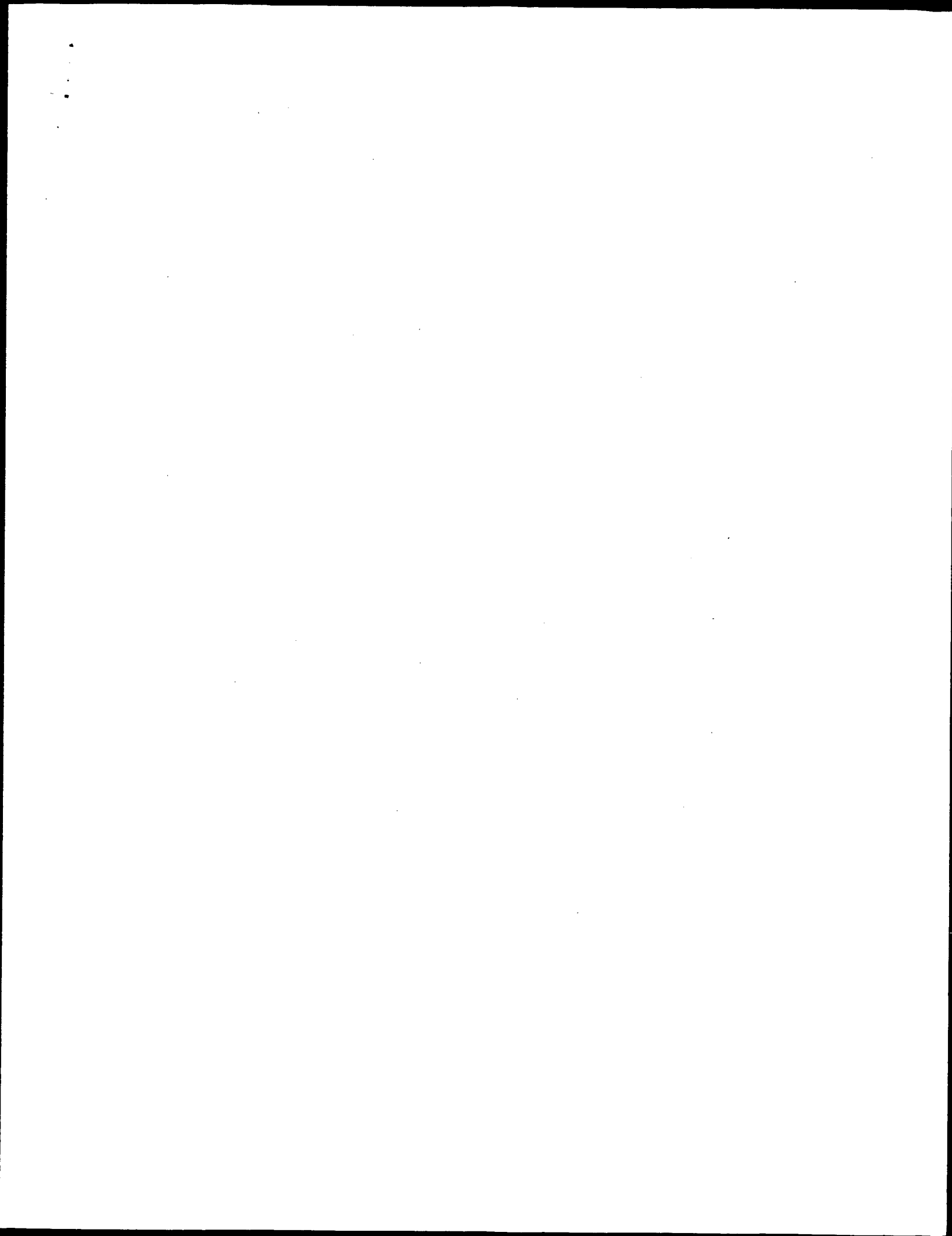
```

407 CCAATACTATGAA.....GAAGTACGTGACCTATGATGTT 444
505 AspleuArgAlaLeuIySglSerValSerAsnLeuGluGlnSerLeuTh 521
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
445 GATCTTAGAAGTATAGAACAGTCTATTAACCAATTAAGAAATCTTTAAC 494
521 rSerLeuSerGluValAlaLeuGlnAsnArgArgGlyLeuAspLeuLeup 538
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
495 TTCCCTGTCCGAGGTGTGCTACAGATAGAAAGGATTAGACTTATTAT 544
538 heLeuArgGluGlyGlyLeuCysAlaAlaLeuIySglucyCysPhe 554
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
545 TCCTTAAAGAGAGGAGCTCTGTGCTGCCCTTAAAGAAATGTTGTTT 594
555 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnIySleuArg 571
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
595 TAGTTGACCATTTCAAGAGTAATCAAGATTCTATGCCCCAAGACTTAGAGA 644
571 slySleuGlu 574
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
645 ACGCCTAGAT 654

```

Mon Feb 25 07:44:30 2002

us-09-171-553b-6.feb22std.rst



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:03 ; Search time 6550.43 Seconds
(without alignments)
4849.227 Million cell updates/sec

Title: US-09-171-553B-9
Perfect score: 2956
Sequence: 1 tgccttttaggttagaac.....aaaaaaaaaaaaaa 2956

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rpd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	472	16.0	549	10	AM657531 110922 MA
2	383.6	13.0	564	11	B1182930 UNC-P-PN-
3	358.6	12.1	602	11	BF712162 MT-P-E6-a
4	293.6	9.9	754	11	A2970955 2M0244H16
5	271.2	9.2	494	11	B1182742 UNC-P-PN-
6	258.6	8.7	701	13	A2351242 1M0089L09
7	255	8.6	774	13	BH122287 RPT-24-2
8	254.4	8.6	591	13	A2430249 1M0214G11
9	253	8.6	632	13	A2814037 2M0081N06
10	246.8	8.3	624	13	A2642336 1M0505M18
11	245.2	8.3	641	13	A2850764 2M0152C18
12	242.6	8.2	493	13	A2349077 1M0085K21

Result No.	Score	Query Match	Length DB	ID	Description
13	240.8	8.1	650	13	A2620796 1M0453112
14	240.6	8.1	625	13	A2813042 2M0080E11
15	234.6	7.9	592	13	A2836232 2M0131E01
16	232	7.8	534	13	A2499035 1M0336P05
17	230.6	7.8	610	13	A2376517 1M0330J20
18	229.4	7.8	644	13	A2356870 1M0098F09
19	228.6	7.7	609	13	A2439189 1M0229N10
20	228	7.7	730	11	B1185535 UNC-P-PN-
21	227.6	7.7	565	13	A2375836 1M0129J20
22	225	7.6	676	13	A2739547 RPT-24-7
23	224	7.6	600	13	A2987182 2M0265J09
24	222.2	7.5	895	10	BE569478 601328494
25	221.6	7.5	597	13	A2406484 1M0175C17
26	221.6	7.5	582	13	BH059027 RPT-24-3
27	220	7.4	576	13	A2730616 RPT-24-1
28	219	7.4	834	13	BH124666 RPT-24-2
29	215.4	7.3	583	13	A2417785 1M0193B08
30	214.2	7.2	669	13	BH040768 RPT-24-2
31	214	7.2	554	11	BF703805 MT-P-O2-a
32	211.2	7.1	659	13	A2741412 RPT-24-7
33	211	7.1	485	11	BF703671 MT-P-E4-a
34	210.6	7.1	704	13	BH051865 RPT-24-2
35	209.6	7.1	461	10	AL120836 DFE2P762N
36	209	7.1	640	13	BH058168 RPT-24-3
37	209	7.1	679	13	A2647666 1M0514J12
38	208.8	7.1	636	13	A2727116 RPT-24-1
39	208	7.0	628	13	A2406003 1M0175N06
40	207.2	7.0	906	11	B1328061 602979695
41	207	7.0	668	13	A2327473 1M0050K14
42	200.6	6.8	631	11	BF702741 MT-P-E3-a
43	198.8	6.7	442	13	A2597961 1M0412K22
44	198.2	6.7	625	13	A2336017 1M0066P05
45	197.8	6.7	605	13	A2416306 1M0191D01

ALIGNMENTS

RESULT 1
LOCUS AM657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos laurus CDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos laurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 549)
AUTHORS Smith,T.P.L., Grose,W.M., Fekling,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pette,G., Holt,I., Karanymcheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 2118013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGATCAGCAGC
Plate: 94 row: E column: 12

Seq primer: ATTTAGGTGACACTATAG.

FEATURES

Source

1. .549

2

11

Library from

swine ovaries

ian follicle

es

CONCLUSION

CDINA

AUTHORS
TITLE
JOURNAL
COMMENT

Caetano, A.
Generation
Library for
Unpublished
Contact: F

R., Johnson
and sequen
om swine ov
d (2001)
omp, D

...R.K. and F
...ce character
...arian folli

SOURCE: House mouse, *Mus musculus*.
ORGANISM: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia, Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus

REFERENCE 1 (bases 1 to 754)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunne@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0244 row: H column: 16
 Seq primer: CGTGTAAACGACGCGCCACT
 Class: plasmid ends
 High quality sequence stop: 754.

FEATURES

Source
 1. 754
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUC2M0244H16"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv: Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g147321149b1AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 226 a 160 c 150 g 218 t
 ORIGIN

Query Match 9.9%; Score 293.6; DB 13; Length 754;
 Best Local Similarity 64.8%; Pred. No. 4.2e-53;
 Matches 474; Conservative 0; Mismatches 244; Indels 14; Gaps 2;
 QY 1882 tggagcgaactactggttaacccctggttccactcagtcctcaacaaccca 1641
 Db 751 tggccttgcaatgacgggtccatcttctgtctatgtctatttattagttccaaa 692
 QY 1642 gatttggtgcatggtcgaactgctcccgagtggtactacatccttgaggagtgctc 1701
 Db 691 gatttctgcattttgggttgcattatcttctgactcctgtaacgtctctttt 632
 QY 1702 ctgtatgaatgactactggtataacgcaaaaagagaacccgtatcccttaccta 1761
 Db 631 tttagcaaatgttagcattcg---gtccgctgagagagagagagccatttactttg 575
 QY 1762 gctgtaatgctggtatgaagcgcgtgtggtgaggaacagagagcgtccctgac 1821
 Db 574 acagttctattagattagagtagcggtgagtaggtacagagacccgtgcttattt 515

QY 1822 acagagaccacagcagcgtctagagaagaactgtgtgactacatgcccactgacagaagat 1881
 Db 514 AAGACCCCAAAATGATA-----TGAAGAACTAGTGCACCTATGATGTTGAT 466
 QY 1882 ctccgaccttaagagagctgtctgaacactgaagagctccctgactcttctgtga 1941
 Db 465 cttagaactcttagacagcttattaccctaattggaagatctttaaacttccctgcga 406
 QY 1942 gtgttctacagaaccggaaggatagatctgtcttcttaagaagaagtgtgtatgt 2001
 Db 405 gtggtgctacagaaatagaagagcattatgactttattttcttaagaagagagactgt 346
 QY 2002 ggcgccttaagaagaatgtgtctctctatgtatgacatcagagacatcagagctc 2061
 Db 345 gctgccccttaagaagaatgtgtttttatgttgaccattcagacataatcaaaagtct 286
 QY 2062 atgaacaagtttgaagaagaatgttagagaagcgtcgaagggaagaagcgtgacccaggg 2121
 Db 285 atggcccaactttagagacagcccttagatcacgtaaaaggaagaaagccacacagga 226
 QY 2122 tggttgaagatgtgtcaacagagctccttgaatgacacaccctgtctgtacg 2181
 Db 225 tggctggaagagcgtgttataatagctccctgctcaccacgtctccacacatagca 166
 QY 2182 gggccctgtagctctgtccctgtactacatggtggcctgtctaatatagttt 2241
 Db 165 ggaacttggatgactcttattgctttgttcttctgttggcccttcgatttataatgta 106
 QY 2242 gtgccttcttagagaacagagtgagtgacgtccagatcatgtgacttaggaacagatc 2301
 Db 105 gttagcttttattttagaagaagatgaatgacagtagacagttatgtactaagcacaacat 46
 QY 2302 caagcctctgt 2313
 Db 45 CGGTCCTTCAG 34

RESULT 5
 B1182742/c 494 bp mRNA EST 10-JUL-2001
 LOCUS B1182742
 DEFINITION UNL-P-FN-do-e-05-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-do-e-05-0-UNL 3', mRNA sequence.
 ACCESSION B1182742
 VERSION B1182742.1 GI:14657151
 KEYWORDS EST.
 SOURCE Sus scrofa
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 494)
 Caetano, A.R., Johnson, R.K. and Pomp, D.
 Generation and sequence characterization of a normalized cDNA
 library from swine ovarian follicles
 Unpublished (2001)
 JOURNAL Contact: Pomp, D
 COMMENT Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: ddunne@unl.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 normalized porcine ovarian follicles library
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source
 1. 494
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection

Query Match	Best Local Similarity	Matches	467; Conservative	8.6%; Score 255; DB 13; Length 774;	60.1%; Pred. No. 8.5e-45;	0; Mismatches 295; Indels 15; Gaps 2;
Db	1437	aaataagcttaccctcactgaagttlccggygaagggagacatgcatagaaagaagctcccc	1496	13	AAACAAGTTGACCTGGCAGCAGTTTCAGGAAGAGGGCTTGTGTGGCCAGGATCTCA	72
Db	1497	atcccacacaccccttgcctatagactgctggtttatgagagccctcgaaatacagta	1556	73	GGATTAAGGGCACTCTGTAAATCAGACCCCAACATCCAGTCATGCAAAAGTGGTCATGA	132
Db	1557	tttagtaccctgtgtaaacagagttgtggtgacatgcatctggttaacccctgtgttc	1616	133	TCGAGTGGCCCCCTGAGACACAGTGTGGGCTTGGCAATCCGGTCTCAGCTCTGTGTGTC	192
Db	1617	caactcagctctcaacccaatccaagaatttgctgcatggtccaaatcgccccagat	1676	193	TATGTCGTGTTTTTAATAGTCCAAAGATTTCGTGATTTTGTCAGCTTAATCTTGAGACT	252
Db	1677	gtactacacatcccttgaggagagtgctcccttgatgaatgatgactacggtataccgaccaa	1736	253	CCTGTATCATGATGATGATACGCCCTTTTATGACAAATTTTGACATCCG--GTCGGCTGGAG	309
Db	1737	aagagaaccggtatcccttaccctagctgtaatgctcgatattaggaacggtgtggt	1796	310	AAGAGAGCCCGGTACTTAACCTTGGCAGTTCTAATGAGATTGAGATAGAGTAGCGCTGAGAT	369
Db	1797	aggaacagggagcagctgcccctgcatcacagaagccagcagcagctagagaagactgtgta	1856	370	AGGTACAGAGAACCGGTGGCTTAATTAAGACCCGCCATTAATGATG-----ACA	417
Db	1857	gtatactatcgccatlgaaagaagatctccagaccttaaggaatgctgttagcaactaga	1916	418	ACTACGTCACACTATGATGTTGATCTTAACACTATTAAGAACACTCTAATACCAAAATTA	477
Db	1917	agagtcctctgactctctgtctgtaagtggtctacagaacccggaagggattagatcgt	1976	478	AGAACTTTACTTCCCTGTCGCCAAGGCGTGTCTACAAATAGAAAGGGGATTGACTTAAT	537
Db	1977	gtttctaaagagaagtggtgattgctgagcagccttaaaagaagaatgtgtcttatgta	2036	538	ATTCTTTAAACAGAGAGACTGTGTGCTGCCCTTAAGAAAGAAATGTTGATTTTAATGTGA	597
Db	2037	tcactcagggagccttagagagctccatgaacaaagcttagaanaaagtttagagaagctcg	2096	598	CCATTACGATTAATCAAGAATTCATAGGCCAAACTTAAGAAACGGCTGATTAATACGTA	657
Db	2097	aaggaagaagaagcttagcagaggggtgtgttttagaagatgattccaagatctccctgat	2156	658	AAGGAAATATGAAAGCCACACAGAGATGTTGCAAACTGGGTAATTAAGTCCCTTGCTT	717
Db	2157	gaaccacctgcttctgctctgaacggggcccttagtactcctgctcctgttactaac	2213	718	CACCACTCTCTCCACCATACGAGACCTTGTATTACTCTTAATGCTTTGCTTAC	774

RESULT	8
AZ430249/c	
LOCUS	AZ430249 591 bp DNA GSS 03-OCT-2000
DEFINITION	M0214G1R Mouse 10kb plasmid UUCGLM library Mus musculus genomic clone UUCGLM0214G11 R, DNA sequence.
ACCESSION	AZ430249
VERSION	AZ430249.1 GI:10554262
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Reilly, 'M.', Rosem,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid insets
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0214 row: 6 column: 11
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 591.

FEATURES	Location/Qualifiers
source	1. .591

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0214G11"
/clone_1id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42mv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g147321141gb/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

Query Match	8.6%	Score 254.4	DB 13	Length 591
Best Local Similarity	66.2%	Pred. No. 1,126	44	
Matches	388	Conservative	0	Mismatches 166; Indels 12; Gaps 1.
QY	1728	ccgacaaaagagagaccgcctatcccttaacctagctgtatgctcgagttgaggaacgac	1787	
Db	589	CCGCGTGGAAAAGAGAACCCGTTACCTTTAACTTTGGCAGAGTTTATTAAGGATTTGGGAGTAC	530	

QY	1788	cgttgcgttaagaaacagagacagctgcgcctcattacaagagcaacagcagcttagaagaaag	1847
Db	529	AGCTGGAATAGTACGAGGAACCGCGCTTAAATTAACACCCGCCAAATCTATGAA-----	473
QY	1848	acttgcgtagacatacagtcgcgcacatgacagaagaatcccgagccctaaaggagctcgttag	1907
Db	474	-----GAGCTACAGTCGAGCTATGGAAGTGTGATTTTGGAAACAAATGAAACAACATCTATAC	422
QY	1908	caacctagaagaagtcoccgacctctcttgcgaagtgttctaaagaaacccgagaggaatt	1967
Db	421	CAAAATTAGAAAGATCTTTAACTTCCCTATCCGAATGAGTGCCTGCAGAAATTAGAGGGGGTT	362
QY	1968	agatcgcctgcttctaaagagaagtglyggtlalygcaagcctlaaaagaagaatgctt	2027
Db	361	AGACTTATTTATTCCTTTAAAGAAAGGAGGAGCTGTCCTCCCTTTAAAGAAAGAAATGTTGATT	302
QY	2028	ctatgtatagatacctagaagacatcagaagaccccttgaaacaaagcttagaanaaagttaga	2087
Db	301	TTATGTTTACCACTTCAGGAGTATATTAAGATTTCTATGGCTAAACTTTAAAGAACCGCTTAA	242
QY	2088	gagcgctcgaagagaaagagagagctctacacaaagggctgttgaaagatagttcaacaglyc	2147
Db	241	TATAGCTAAAGAGCAAAAGAGAAAGCCCAAGAGATGGTTTGAAACCTGGTTTAAATAAGTC	182
QY	2148	tccttgcgttagacaccccgctcttcctgcgccttagacggggcccctcagctagctcctcctgct	2207
Db	181	CCCTTGGCTCACACACTCTCTCTCCACTATAGCAAGAACCTTTAATTTACACTTATGCTTTT	122
QY	2208	acttaacagttgggcctctgcttaattaaatgaattgttgccttgccttgccttgccttgccttgcct	2267
Db	121	GCTTACTTTTGGCCCATGGATCCTTAATTAAGATTAGCTTTTAAATAGAGAAAGGATATA	62
QY	2268	tgcaagtcacagatacaatgctacttagagcaacagatccaagaagcctctgcg	2313
Db	61	CGCACTCCAAAGTTATGTGTACTTAAGCAACAAATATGAGGGCTCTCTCAAG	16

TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunne@genetics.utah.edu Insert length: 10000 Std Error: 0.00 Plate: 0081 row: N column: 06 Seq primer: CACACAGCAACACGCTATGACC Class: plasmid ends High quality sequence stop: 632. Location/Qualifiers 1. 632
FEATURES	
source	

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M081N06"
/clone_id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1Arl29072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      203 a      115 c      138 g      176 t
ORIGIN

```

```

Query Match      8.6%; Score 253; DB 13; Length 632;
Best Local Similarity 66.6%; Pred. No. 2.4e-44;
Matches 383; Conservative 0; Mismatches 180; Indels 12; Gaps 1;

```

```

QY 1737 aagagaccggtatccctaccctagctgtaatgctggaatgagagcgccgtgagcgt 1796
DB 2 AAGAGAACCCGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 61
QY 1797 aagagagggagacgtccctgacacagagacacagacagacagacagacagacagacag 1856
DB 62 AGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 110
QY 1857 gctacatgagcagacagacagacagacagacagacagacagacagacagacagacagac 1916
DB 111 -CTACGTCAGACGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 169
QY 1917 aagatccctaccctctctgctgctgctgctgctgctgctgctgctgctgctgctgct 1976
DB 170 AGAATCTTTAACTTCCCTGCTCCGAAAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 229
QY 1977 gttctaaagagagtggtatgtagacgtttaaagaagaatggttgcctctatgtaga 2036
DB 230 ATTCTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
QY 2037 tcaactcagagacacagacacagacacagacacagacacagacacagacacagacacag 2096
DB 290 CCAATTCAGAGATATCAAAATCTATGCGCAAACTTAGAGACGCTATACGTAA 349
QY 2097 aagggagagagagagagagagagagagagagagagagagagagagagagagagagag 2156
DB 350 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409
QY 2157 gaccacccgtcttgcctgacggggccctcagtagcgcgtcgtcttacttaacagt 2216
DB 410 CACCACTCTCTCCACCACTACAGAGACCTTTAATTAATTAATTAATTAATTAATTAATTA 469
QY 2217 tgggacctgtaataataggttgccttgccttgccttgccttgccttgccttgccttgcct 2276
DB 470 TGGCTCTCTGATCTTAAATAGTATGCTTTTATTAAGAGAGAGATTAATTAATTAATTA 529
QY 2277 gatcatggtacttaggaacagatcacaagccttc 2311
DB 530 GGTATGCTACTAAGCAGCAATATCGGGCTTC 564

```

```

RESULT 10
A2642336 624 bp DNA GSS 14-DEC-2000
LOCUS A2642336
DEFINITION 1M0505M18F Mouse 10kb plasmid U06C1M library Mus musculus genomic
clone U06C1M0505M18 F, DNA sequence.
ACCESSION A2642336
VERSION A2642336.1 GI:11768843
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 624)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contract: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0505 row: M column: 18
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 624.

```

```

FEATURES
source
1. 624
location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0505M18"
/clone_id="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1Arl29072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      203 a      116 c      131 g      174 t
ORIGIN

```

```

Query Match      8.3%; Score 246.8; DB 13; Length 624;
Best Local Similarity 66.1%; Pred. No. 5.2e-43;
Matches 378; Conservative 0; Mismatches 182; Indels 12; Gaps 1;
QY 1742 aaccggtatccctaccctagctgtaatgctgagagagcgccgtgagcgtgagaa 1801
DB 1 AACCGTTACCTTAACTTGGCAGTTCATTAAGATTGGAGTAGCAGCTGGAGTAGTA 60

```



```
/db_xref="taxon:10090"
```

/clone="UUCG1M0453112"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 180 a 131 c 128 g 211 t
 ORIGIN

Query Match 8.1%; Score 240.8; DB 13; Length 650;
 Best Local Similarity 67.0%; Pred. No. 1e-41;
 Matches 363; Conservative 0; Mismatches 167; Indels 12; Gaps 1;

QY 1772 tcgattagaggaagccggttggtgtagaagagagagagccttcgtagacagagacc 1831
 Db 649 TAGATTGGGAGTAGACAGCTGAGTACGAGAAATCGCTGCTTAATTAAGACCCCC 590
 QY 1832 agcagctagagaagagacttgtagatcatatgcgacatgacagaagatcgcagcct 1891
 Db 589 CAATTCATA-----AGACTACGTCGACGTATGATATGATATGATTAAGACTA 542
 QY 1892 taagagagctgttagcaacctagaagatcctcgtactcttctgtagagtggtttac 1951
 Db 541 TAGAAGAGTCTATACCAAAATTAAGAAATCTTAACTCCCTGCGAAGTGTGCTGC 482
 QY 1992 agaaccgagagagtagatcgtctgtttcctaagaaggttggtttagtgcacccctaa 2011
 Db 481 AAAATPAGAGGGATTTAGACTTATTCTTTAAAGAGAGAGACTCTGCTGCTTTAA 422
 QY 2012 aagaagaatgttctatgtatgatacactcagagagccatcagagactcagagaacg 2071
 Db 421 AAGAAGATGTTGTTTATATGTTGACCATTCAGAGATTAATCAAGATTCTATGCGCAAC 362
 QY 2072 ttagaagaattagagagagcgtcgaagaggaagagagcgtgaccagaggtgtgttag 2131
 Db 361 TTAGAGAAAGGCTAGATTTAGCTAAAGAGAAAGAAAGCAAGAGATGTTTAA 302
 QY 2132 gatggttaacagagctcctccttgatgacacccctgcttctcctcgaagggccctag 2191
 Db 301 GCTGTTTAAATAGTCCCTTGCTGACACACTCTCTCTCCACATATAGCAAGACCTTAA 242
 QY 2192 tagtctgctcctgttacttaacagttgggctgtccttaataagttgtgtcctttg 2251
 Db 241 TTACACTTATGCTTTGCTTCTTATGCTTGGCCATGCATCTTAAATAGTTAGTACTTTTA 182
 QY 2252 ttaagaagagagtgtagtgcagtcagtcagtcagtcagtcagtcagtcagtcagtc 2311
 Db 181 TTAGAGAAAGATTAAGCAAGTCAAGTTATGTTAGTAAAGCAACATATATGCGTCTTC 122
 QY 2312 tg 2313
 Db 121 AG 120

RESULT 14

AZ813042/c
 LOCUS AZ813042 625 bp DNA GSS 20-FEB-2001
 DEFINITION 2M0080E11F Mouse 10kb plasmid UUCG1M library Mus musculus genomic clone UUCG2M0080E11 F, DNA sequence.

ACCESSION AZ813042
 VERSION AZ813042.1 GI:12982881
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0080 row: E column: 11
 Seq primer: CTTTAAAGCAGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 625.
 Location/Qualifiers
 1..625
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG2M0080E11"
 /clone_lib="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 191 a 128 c 120 g 186 t
 ORIGIN

Query Match 8.1%; Score 240.6; DB 13; Length 625;
 Best Local Similarity 63.8%; Pred. No. 1.1e-41;
 Matches 404; Conservative 0; Mismatches 214; Indels 15; Gaps 2;

QY 1574 acagtggtggtgagatgatactggtggtggtggtggtggtggtggtggtggtggt 1633
 Db 618 ACACAGGTGAGCTTGCATTAACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559
 QY 1634 aatccaaagattgtgtgtatgtgtccaaatgctcccccagagtgatcactacatcgtgag 1693

|||||||
Db 562 TGCAGTCCAAGTC 574

Search completed: February 23, 2002, 23:37:18
Job time: 9439 sec

OM of: US-09-171-553b-10 to: EST.* out_format : pfs

Date: Feb 24, 2002 8:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+g2n.model -DEV=x1h
-q/cgn2.1/USPTO_spoil/US09171553/runat_22022002.122825.11137/app_query.fasta.1.3312
-bb=EST -Qfmt=fastap -SUFFIX=13522std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-EGAPOP=6.000 -EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELPO=6.000 -DELEXT=7.000 -START=1 -MATRIX=bls6sum62
-TRANS=human4.0.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
-NORM-ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09171553_@CGN1.1.9008 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-171-553b-10
Query length: 657
Database: EST.*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 3515.700000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:AM657531	+	888.00	1216.45	1.4e-58	549
gb_gss:AZ270955	-	875.50	1196.21	1.8e-57	734
gb_gss:BH059027	+	859.50	1172.93	3.7e-56	882
gb_gss:BH122827	+	843.50	1151.93	5.3e-55	774
gb_gss:AZ251442	-	814.00	1112.30	8.6e-53	701
gb_est1:B1151181	+	787.00	1073.36	1.3e-50	845
gb_gss:AZ2850764	-	770.50	1053.32	1.7e-49	641
gb_gss:AZ2813042	-	757.50	1035.27	1.6e-48	625
gb_est1:B1106669	+	757.00	1031.27	2.8e-48	921
gb_gss:AZ236870	-	734.50	1003.74	9.6e-47	644
gb_gss:AZ2376517	-	730.50	998.76	1.8e-46	610
gb_gss:AZ2430249	-	729.00	997.00	2.3e-46	591
gb_gss:AZ242336	+	721.50	986.16	9.1e-46	624
gb_gss:AZ2379937	+	721.50	985.44	1.0e-45	673
gb_gss:AZ214037	+	713.50	975.04	3.8e-45	632
gb_gss:AZ214142	+	711.00	971.07	6.3e-45	659
gb_gss:AZ237473	-	696.50	952.72	6.7e-44	565
gb_gss:AZ2375836	+	693.00	948.45	1.2e-43	534
gb_gss:AZ2499035	-	669.00	946.19	1.5e-43	676
gb_gss:AZ217785	-	665.50	937.29	4.8e-43	583
gb_est1:B1109841	+	681.50	928.91	1.4e-42	787
gb_gss:AZ2620796	+	671.50	916.99	6.5e-42	650
gb_gss:AZ2836232	+	670.00	915.82	7.6e-42	592
gb_est1:BF608155	+	667.00	912.54	1.2e-41	833
gb_est1:BF569478	+	666.00	910.48	1.5e-41	895
gb_gss:AZ216306	-	662.50	905.29	2.9e-41	605
gb_est1:BF159914	+	659.00	899.36	6.2e-41	680
gb_gss:BH054319	-	653.50	890.50	1.9e-40	724
gb_gss:AZ236017	-	652.50	891.22	1.8e-40	625
gb_est1:BF181106	+	645.00	876.70	1.0e-39	969
gb_gss:AZ2764389	+	644.50	877.72	1.0e-39	609
gb_gss:AZ2647666	+	638.00	870.48	2.5e-39	679
gb_est1:B1106591	+	628.50	853.99	2.1e-38	728
gb_est1:B1158931	+	624.50	852.24	2.5e-38	716
gb_gss:AZ289131	+	624.50	852.24	2.6e-38	656
gb_gss:AZ249077	+	618.00	842.01	9.8e-38	750
gb_est1:B1082444	-	612.00	835.08	2.3e-37	644
gb_gss:AZ2960032	-	611.50	834.02	2.7e-37	675

```
gb_gss:AZ296293 - 603.50 824.77 8.9e-37 566 | AZ296293 2M0282L19F Mouse
gb_est1:BF786434 - 602.00 819.37 1.8e-36 801 | BF786434 602112966F1 NCI.C
gb_gss:AZ2897402 - 599.00 818.42 2.0e-36 575 | AZ2897402 RPT-24-176N10.TJ
gb_gss:BH055839 + 596.00 813.57 3.7e-36 620 | BH055839 RPT-24-93P16.TVB
gb_gss:AZ237496 - 596.00 813.08 4.0e-36 653 | AZ237496 1M0496123R Mouse
```

seq_name: gb_est1:AM657531

seq_documentation block:

```
LOCUS AM657531 549 bp mRNA EST 25-APR-2001
DEFINITION 110922 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AM657531
VERSION AM657531.1 GI:7423429
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 549)
```

REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahnensterg,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Kohrer,G.A., Chitko-McKown,C.G., Perlee,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.

AUTHORS

Sequence evaluation of four pooled-tissue normalized bovine CDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

CONTACT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -mnscore 18
and -mismatch 12 options.

PCR PRIMERS

FORWARD: AGGAACCACTATGACCAT
BACKWARD: GTTTCACACTCAGCAGC
Plate: 94 row: E column: 12
Seq primer: ATTATGACACTATGAC.

FEATURES

source

```
1..549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site:1: XbaI; Site:2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 145 a 144 c 121 g 139 t
ORIGIN
```

alignment_scores:

```
Quality: 888.00 Length: 183
Ratio: 1 Gaps: 1
Percent Similarity: 95.078 Percent Identity: 89.617
```

alignment_block:

US-09-171-553b-10 x AM657531 ..

Align seg 1/1 to: AM657531 from: 1 to: 549

290 Set-ProGlyValProValIysThrGlyGlnArgLeuPheSerLeuIleG1 306

3 TCCCCCGGGTTCCTTAAGACAGACAGACAGCTTCAGCTCATCA 52

306 nGlyAlaIleGlnAlaIleAsnSerThrProAspAlaIleIleSerC 323

529 LeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGlyGlyLe 545
 683 TTTCAAAACAGAAAGAGTTTATGATTTATCTTTTAAAGGAAGGTGACT 732
 545 uCysAlaAlaLeuLysGlyLysCysCysPheTyrValAspHisSerGlyA 562
 733 ATGTGACGCTTAAAGAGAAATGCTGCTTATGACGACACAGCGGAA 782
 562 laileArgAspSerMetAsnLysLeuArgLysLeuGlnArgArg 578
 783 TAGTCAGAGATTTCTATGCAAAATTGAGGAAAGACTGGAAGGGA... 829
 579 ArgGluArgGluAlaAspGlnGlyTyrPheGlnGlyTyrPheAsnArgSe 595
 830TGGTTCAGAGGTGTGTACTCTGTAATC 855
 595 rProTyrMetThrThrLeuLeuSer 603
 856 TCCATGATGACCACTTACTTTCT 880
 seq_name: gb_gss:BH122287

Seq documentation block: 774 bp DNA GSS 19-JUL-2001
 LOCUS BH122287
 DEFINITION RPI-24-288P13.TV RPI-24 Mus musculus genomic clone RPI-24-288P13
 , DNA sequence.
 ACCESSION BH122287
 VERSION BH122287.1 GI:14965799
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 774)
 Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M.,
 Tesegay, G., Geer, K., Krol, M., Shvartsbeyn, A., Geuregeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 Other GSSs: RPI-24-288P13.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@igrr.org

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Clones are derived from the mouse BAC library RPI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 288 row: P column: 13
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers

FEATURES
 source
 1..774
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-24-288P13"
 /clone_id="RPI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1. Site 1: BamHI. Site 2: BamHI.
 RPI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 BASE COUNT 221 a 162 c 172 g 219 t
 ORIGIN

alignment_scores:
 Quality: 843.50 Length: 262
 Ratio: 3.749 Gaps: 2
 Percent Similarity: 85.878 Percent Identity: 61.832

alignment_block:
 us-09-171-553b-10 x BH122287 ..

Align seg 1/1 to: BH122287 from: 1 to: 774

355 TrrpLysArgAsnLysLeuThrLeuThrGluValSerGlyLysGly 371
 2 TTTGGGGGGAACAAAGATTGACTCTGGCAGACCTTCAGAGAGAGGCT 51
 371 rCysAlaGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThr 388
 52 TTTGTTGGGCCGACCTCAGATTAAGGACACCTCTGTAATCAGACC 101
 388 alValTyrGluGlnAlaSerGlnAsnGlnTyrLeuValProGlyTyrAsn 404
 102 AGAACATCCAGCTCAGCAAAAGTGTGATCAGTACGTGCCCCCTAGAC 151
 405 ArgTrrpAlaCysAsnThrGlyLeuThrProCysValSerThrSerVa 421
 152 ACAGTGTGGGCTTCAATACCGGTCTCACCCTTGTGTGTACTGTCTGT 201
 421 lPheAsnGlnSerLysAspLeuCysValMetValGlnIleValProArg 438
 202 TTTAATAGTTCACAAAGATTTCTGATTTGGTTCACCTTATTCCTAGAC 251
 438 alTyrTyrHisProGluGluValValLeuAspGlnTyrAspTyrArgTyr 454
 252 TCTGTATCATGATGATGAGCGCTTTTACACAAATTTGACATCGGCTG 301
 455 AsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuG 471
 302 ...CGCTGAGAAAGAGAGCCCGTTACCTTACCTTGGCAGCTTCATTAAG 348
 471 yLeuGlyThrAlaValAlaGlyThrGlyThrAlaAlaIleThrG 488
 349 ATTGAGAGTACGCGCTGAGTAGTACAGAACCGCTTAAATTAAGA 398
 488 lProGlnGlnLeuGluLysGlyLeuGlyLysLeuHisAlaAlaMetThr 504
 399 CCCCCCAATACATAGAA.....CACTACGCTGACGCTATGAT 436
 505 GlnAspLeuArgAlaLeuLysGlnSerValSerAsnLeuGlnLysSer 521
 437 GTTGATCTTGAACATATGAAACAGTCTATTAACCAATTAAGAAATCTTT 486
 521 urhSerLeuSerGlnValValLeuGlnAsnArgArgGlyLeuAspLeu 538
 487 AACTTCCTCTCCGAAGTGTGCTACCAAAATTAAGAGGATTTGACTTAT 536
 538 eupheLeuArgGluGlyLysLeuCysAlaAlaLeuLysGlnLysCys 554
 537 TATTCCTTAAAGAACGAGACTCTGCTCCCTTAAAGAAAGAAATGTGA 586
 555 PheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeu 571
 587 TTTATCTTGAACATTCATGATTAAGATTAAGATTAAGATTAAGATTAAG 636
 571 glyLysLeuGluLysArgArgArgGluArgGluAlaAspGlnGlyTrrp 588
 637 AGAAGCGCTAGATATACGTAAGAGAAATTAAGAAACCAAGATTCGT 686
 588 heGluGlyTrrpPheAsnArgSerProTrrpMetThrThrLeuLeuSerAla 604
 687 TCGAAAGCTGGTTAATTAAGTCCCTTGGCTACACCTCTTCTCCAC 736
 605 LeuThrGlyProLeuValValLeuLeuLeuLeu 616

737 ATAGCAGACCTTGTACTCTTATGCTTTGCTT 772

seq_name: gb_gss:AZ351242

seq_documentation_block:

LOCUS AZ351242 701 bp DNA GSS 29-SEP-2000
 DEFINITION 1M0089L09F Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 clone UNGC1M0089L09 F, DNA sequence.

ACCESSION AZ351242 GI:10430479

VERSION AZ351242.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 701)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relilly,
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.
 and Wright, D., Weiss, R.

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0089 row: 1, column: 09
 Seq primer: CGTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 701.

Location/Qualifiers
 1..701

FEATURES

source

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNG1M0089L09"
 /clone_lib="Mouse 10kb plasmid UNGC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: PMD42ntv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g1473214(g1473214)29072.1) a copy-number
 inducible derivative of plasmid RL. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 211 a 149 c 134 g 206 t 1 others

ORIGIN

alignment_scores:
 Quality: 814.00 Length: 234
 Ratio: 3.895 Gaps: 4
 Percent Similarity: 89.316 Percent Identity: 67.521

alignment_block:

US-09-171-553B-10 x AZ351242/rev ..

Align seg 1/1 to reverse of: AZ351242 from: 1 to: 701

```

417 ValSerThrSerValPheAsnGlnSerLysAspLeuCyValMetValGI 433
||||| ||||||| ||||||| ||||||| ||||||| |||||||
700 GTGTCTATGCTGCTGCTTTTAAATGCTCCAAAGATTTGCAATTTGGTTCA 651
433 nileValProArgValTyrTyrHisProGlnGluValValLeuAspLut 450
||||| ||||||| ||||||| ||||||| ||||||| |||||||
650 ACTTATTCCTAGACTCCTGATTCATGATGATAGCTCTTTTACACNAT 601
450 YRAAPYRATGTYRAsnArgProLysArgLuproValSerLeuThrLe 466
||||| ||||||| ||||||| ||||||| ||||||| |||||||
600 TTGACATCGGGCTC...CGCTGGAGAAAGAACACCGTTACCCCTAATCTT 554
466 uAlaValMetLeuGlnGlyLeuGlyThrAlaValAlGlyValGlyThra 483
||||| ||||||| ||||||| ||||||| ||||||| |||||||
553 GCAGTTCTATTAGGATGGAGATGAGCGGTTGGAGTAGGTAACAGAACTG 504
483 lAlaLeuLeuLeuThrGlyProGlnGlnLeuGlnLysGlyLeuGlu 499
||||| ||||||| ||||||| ||||||| ||||||| |||||||
503 CTGCCTTATTATTAAGACCCCAATACTATGAA.....GAACCTA 466
500 HisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlnSerValSer 516
||||| ||||||| ||||||| ||||||| ||||||| |||||||
465 CGTCGAGCTATGATGATGTTGATCTTAGAATCTATAGACAGCTATTAACAA 416
516 nLeuGlnGlnSerLeuThrSerLeuSerGlnValAlValLeuGlnAsnArg 533
||||| ||||||| ||||||| ||||||| ||||||| |||||||
415 ATTAGAAAGAACTTAACTTCCCTGCGAAGTGCTGCTACAGATATGAA 366
533 rGgLYLeuAspLeuLeuPheLeuArgGlnGlyGlyLeuCyAlaAlaLeu 549
||||| ||||||| ||||||| ||||||| ||||||| |||||||
365 GGGATTAGACTTATTATCTCTTAA...GGAGGACTGTGCTGCCCTA 319
550 LysGlnGlnCyScyPheTyrValAspHisSerGlyAlaLeuArgAspSe 566
||||| ||||||| ||||||| ||||||| ||||||| |||||||
318 AAAGAGAGGTGTTGTTTATGTGACCATTCAGAGTAATCAAGATTC 269
566 rMetAsnLysLeuArgGlyLysLeuGlnArgArgArgGlnArgGlu 583
||||| ||||||| ||||||| ||||||| ||||||| |||||||
268 TATGCCCAACTTAGAGAACCCCTAGATATGCAAAAGAGAGAA 219
583 lAspGlnGlyThrPheGlnGlyThrPheAsnArgSerProTrpMetThr 599
||||| ||||||| ||||||| ||||||| ||||||| |||||||
218 GCCAACAGAGATGTTGCAAGCTGTTTAAATAGTCCCTTGGCTCAC 169
600 ThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeu 616
||||| ||||||| ||||||| ||||||| ||||||| |||||||
168 ACTCTCCCTCCACCATAGACGACCTTAAATACACTTATGCTTCT 119
616 uThrValGlyProCysLeuLeuAsnArgPheValAlaPheValArgGua 633
||||| ||||||| ||||||| ||||||| ||||||| |||||||
118 TACTTTGGCCCTGCACTTAAATAGCTTATGATCTTTATTTAGAGAA 69
633 rGValSerAlaValGlnLeuMetValLeuArgGlnGlnGlyLeu 649
||||| ||||||| ||||||| ||||||| ||||||| |||||||
68 GGATTAATGCACTACAGTTATGTTAGTACGCAACATATCGGCTCTT 19
602916935F1 NCL_CGNP_Lu29 Mus musculus cDNA clone IMAGE:5067273 5',

```

seq_name: gb_est2:B151181

seq_documentation_block:

LOCUS B151181 845 bp mRNA EST 05-JUL-2001
 DEFINITION 602916935F1 NCL_CGNP_Lu29 Mus musculus cDNA clone IMAGE:5067273 5',
 mRNA sequence.

ACCESSION B151181 GI:14611182

VERSION B151181.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 845)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.nih.gov>
 Plate: LAM1181 row: f column: 10
 High quality sequence stop: 774.

FEATURES
 source
 1. 845
 /organism="Mus musculus"
 /strain="C57BL/6J (f1)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5067273"
 /clone_id="MCI CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /stem_cell_origin=""
 /lab_host="DH10B"
 /note="Organ: lung; Vector: PCMV-SPORE; Site: 1: Salt; Site: 2: Not; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 225 a 188 c 216 g 215 t 1 others
 ORIGIN

alignment_scores:
 Quality: 787.00 Length: 288
 Ratio: 3.422 Gaps: 8
 Percent Similarity: 79.861 Percent Identity: 58.333

alignment_block:
 US-09-171-553B-10 x B1151181 ..

Align seg 1/1 to: B1151181 from: 1 to: 845

```

309 pheglnalaleasnserthaspproaspalahtserseystrip 325
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 TTTTATGCCCTTAACAGACTGATCCAGCGCTACTGAGCGCTGCTGCT 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
325 ucysleusersegiypropotyrytyrlygluylmetalaalysgluarg 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 ATGCCCTGCTTGGGTCGCCCTTATATGAGGAATGCGCTTCATGAG 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 yspheasnvalthrlysluhisargasnincyshttrpglyserarg 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ATTTCACAGAACAGACGAGCCAT...ACTTCTCTCTCTGGGATACAGA 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 asnlseuthrleuthrlygluvalserglylys.....glythrcys 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 CAAAAGACGACCTGAGTATCCGCGAGGAATCCAGGCTCTCTAT 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 egllyalalaproproserhisglnhisleucytsrsethrvalval 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 AGGTACCCCACTTCCACTCACAACACCTATGCGCAATTCAGTCCG 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
390 yrglglinalasergluasnglntyrleuvalproglytyrasnarg 406
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 TGTCACAGACGAGACTAATTACTTGTACTTCCCGGTTGGATGG 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 TTPALACysantrglyleuthrprocyvalserthservalpheas 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 TGGGCTTGACATACAGACTTACTCCCTGTGTATCAACTAAGGTTTAA 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 nglnserlyaspleucysvalmetvalglnilevalproargvaltyr 440

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 TTCACTTCATCATGATT.TGTGTCATGATCCAGCTCTTACCCCGGTATATT 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 YTHISPROGLUGLUVAlValleuaspglutyraspyrargtyrasnarg 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
398 ATCACTTCATCCAGTTTAA.....GAGAGAACTATCTGCTGCGCGCG 441
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 PROYSARGGUPROVALSERLEUTHRLEUAlValmetleuengl 473
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 TCAAAAGAGAACCAATTACTTACCTTGCTGCTGATTCAGGGAATAGG 491
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 YTHAlAlAlGlyValGlyThrlythralalaleuilethrglypro 490
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 TATGACAGTAGAGACTGGGACGAGAGTGCAGCTTGATGAGGAAGAC 541
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 lnglnleuylglyglyleuylgluleuhsalalaleuilethrgluasp 506
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 AG.....GGAATTCAGTCTTGAAGGATGCTGCAATGAAGAC 579
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 leuarglaleuylsgluservalserasnleuylgluileuthrse 523
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
580 CTAGCGCAATAGAGAACTGATGACGCTTTAGAAAATCTTGGACCTC 629
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 rleusergluvalValleuaglnasnargargly. leuasp leuaphe 539
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
630 CCGTCTGAGGTAGTTTACAGAACAGAGAGCTTGCTGCTTCTGCTTC 679
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 leuarglulglyleucysalalaleuylglu. gluycysgphet 556
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
680 CTACAGAGAGAGAGAGCTGCTGCTGCTTAAAGAAAGAGCTGCTCT 729
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
556 yvalasphisserglyalalearg. aspsrmetasnlyseuargly 572
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
730 ATGCAAGTCAATACAGAGATAGTGAAGAGCTATCCGAACTGGAGAG 779
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 slvs...leuyluargargargarglyuarglualaspglnglytrp 588
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
780 AAGCATTCGAGGAGGAGACCGGAGGAGGAGCTGCTACGCGGGTGGT 829
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 heglugly 590
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
830 AGTCCCGG 837
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_gss:A2850764
seq_documentation_block: 641 bp DNA GSS 21-FEB-2001
LOCUS A2850764
DEFINITION 2M0152C18R Mouse 10kb plasmid U06C1M library Mus musculus genomic
ACCESSION A2850764
VERSION A2850764.1 GI:13036088
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 641)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

```


musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 191 a 128 c 120 g 186 t
ORIGIN

alignment_scores:
Quality: 757.50 Length: 207
Ratio: 4.051 Gaps: 2
Percent Similarity: 90.338 Percent Identity: 69.082

alignment_block:
US-09-171-553B-10 x A2813042/rev ..

Align seg 1/1 to reverse of: A2813042 from: 1 to: 625

```

408 Alacysasnrhrglyleuthrprocyvalserthrservalpheasngl 424
|||||.....|
607 GCTTCACATTAACGGCTCTACTCTGTGTGCTATGCTGTTTAAATAG 558
|||||.....|
424 nserlyasplencysvalmetvalglnilevalproarvaltyrtyrh 441
|||||.....|
557 TTCTAAAGATTTCGACATTTGGTTCAGCTTATCCAGACCTCTGATTC 508
|||||.....|
441 isproglugluvalleuaspluftyraspytyrtyrasnarypro 457
|||||.....|
507 ATGATGATAGCTCTTTTATGATTAATTTGAACATCGGGTC...CGCTGG 461
|||||.....|
458 Lysargluprovalserleuthrleualalmetleuglyleuglyth 474
|||||.....|
460 AAAAGAAACCCGTTACCTTAACCTTGGCAGTCTATTAGGATGGAGT 411
|||||.....|
474 talavalglyvalglythrglythralalaleuilethrglyproglng 491
|||||.....|
410 AGCAGCTGAGTAGGTACAGAAACCGCTGCTTATTAAGACCCCAAT 361
|||||.....|
491 lnleuglyuglyleuglyleuileuileuileuileuileuileuileu 507
|||||.....|
360 ACTATGAA.....GAACTACGTGACCTATGATTTATGATCTT 323
|||||.....|
508 Argalaleuylusgluservalseranleugluiserleuthrserle 524
|||||.....|
322 AGAATATATACAGCTCTATACCAATTAAGAAATCTTAACTTCCTCT 273
|||||.....|
524 usergluvalleuileuglnasnargrglyleuasplleuileuileu 541
|||||.....|
272 GTCTGAAGTGTCTGCTCAAAATAGAGGGGATTAGACTTATATCTCTTA 223
|||||.....|
541 rggluglyleuglyleucysalaleuileuglyleucyscyspetyrval 557
|||||.....|
222 AAGAGAGAGACACTGTGCTTAAAGAAATGTTGTTTATGTTT 173
|||||.....|
558 ASPHISerglyalalleargaspermetasnlyseuileuileuileu 574
|||||.....|
172 GACCATTCAGAGATATCAAGATTTCTATGCTTAAAGGAGGAGGCTT 123
|||||.....|
574 ugluarargargargluarglualeuileuileuileuileuileuileu 591
|||||.....|

```

```

122 AGATATACGTTAAAGAGAAAGAGAGAGCAACCAAGATGCTCAAAAGCT 73
|||||.....|
591 rppheasnarserprotrpmetthrthrleuileuileuileuileuileu 607
|||||.....|
72 GGTTTAAATAAAGTCCCTGGCTCACACATCTCTCTCCACTATAGCAGCA 23
|||||.....|
608 Proleuvalleuileuileu 614
|||||.....|
22 CCTTATATTAACCTATGCTT 2
|||||.....|

```

seq_name: gb_estc2:BI106669
seq_documentation_block:
LOCUS BI106669 921 bp mRNA EST 26-JUN-2001
DEFINITION 602893272F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5038336 5',
RNA sequence.
ACCESSION BI106669
VERSION BI106669.1 GI:14557562
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 921)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LAM1105 row: P column: 17
High quality sequence stop: 805.
Location/Qualifiers
1. 921
/organism="Mus musculus"
/strain="C2EHC II (Feral)"
/db_xref="taxon:10090"
/clone="IMAGE:5038336"
/clone_lib="NCI CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 249 a 211 c 229 g 232 t
ORIGIN

alignment_scores:
Quality: 757.00 Length: 304
Ratio: 3.181 Gaps: 10
Percent Similarity: 78.289 Percent Identity: 53.947

alignment_block:
US-09-171-553B-10 x BI106669 ..

Align seg 1/1 to: BI106669 from: 1 to: 921

```

299 Glnargleupheserleuileglnlyalaphelinalaleasnerth 315
|||||.....|
4 CAGAGGATGTTTAACTAGTACAGAGAGCGCTTATATCCCTTAACACAC 53
|||||.....|
315 rAspproaspalarhrsersercystripheucysleuSerSerclyProp 332
|||||.....|

```

```

54  TGATCCAGCGCTACTGAGGACTGCTGGCTATGCTGCTGGGCTCCG 103
332  rcttTyrGluclyMetAlaLysGluArgLysPheAsnValThrlySglu 348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104  cttATTATAGAGAAATCGCTTCATGGAATTTCAACAGAACCGACG 153
349  HisArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrG1 365
      ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
154  CAT...ACTTCCTGCTCTGGGGTACAGACAAAAATGACCCCTGACGTA 200
365  vtaISerGlyLys.....GlyThrCysLeuGlyLysAlaPropSerH 380
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  ACTATCCGAGAGAAATCCAGCTCTGTATAGATCCCAACCTTCACATC 250
380  iAGlnHisLeuCysTyrSerThrValValTyrGluGlnAlaSerGluAsn 396
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
251  ACAAACACCTATGCGACAAATTCAGTCCGTCCAGAACGGAACCTGAT 300
397  GlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGlyLe 413
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
301  TACTATCTGTACCTTCCCGGTTGAGTGGCTGCTGCAATACAGACT 350
413  unhrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysV 430
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
351  TACTCCCTGTGTATCACTAAAGTTTAAATTCACTCATGATTTTCGTG 400
430  aIMetValGlnIleValProArgValTyrTyrHisProGluGlnValAla 446
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
401  TATATATCCAGCTGTACCCCTGTATATTTATACACCTGCATCCAGTTTA 450
447  LeuAspGlyTyrAspTyrArgTyrAsnArgProLysArgLysProValSe 463
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
451  GAAGACAGCTATGCGGCGCG.....CGGTCAAAAAGAAACCAATTAC 494
463  rleuThrIleuAlaValMetLeuGlyLeuGlyThrAlaValGlyValGlyT 480
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
495  TTTAAACCTCGCTGCTCATGGAATAGATAGCAGTAGAGAGCGGA 544
480  hTgIlyThrAlaAlaLeu..Ile..ThrGlyProGlnGlnLeuGlyLysGlyL 496
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
545  CGGAGAGTGCAGCTTTCGATAGAACGGAAGACAG.....GGAA 582
496  euGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 512
      :: :|||:|||||:|||||:|||||:|||||:|||||:|||||:
583  TTCAGTCTTTGAGGATGCTGTCATGAAGACCTAGCGCAATAGAGAG 632
513  SerValSerAsnLeuGlnGluSerLeuThrSerLeuSerGluValAlaLe 529
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
633  TCCATTGACGCTTTAGAAAAAATCTTGACCTCCCTGCTGAGGTAGTTT 682
529  uGlnAsnArgArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuG 546
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
683  ACGAACAAGAGAGAGCTTTGATTCGTCCTCAAAAGACAGAGCACTGTT 732
546  ysaAlaAlaLeuLysGluLysCysCysPheTyr..ValAspHisSerGly..A 562
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
733  GTGCTGCCCTTAAGAAGAGTCTGCTCTATTGACATCAACACAGAAAT 782
562  laIlaIArgAspSer..MetAsnLysLeuArgLysLysLeuGluArg.. 577
      ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
783  TAGTTAGAGACTCTATGACAGAACTTGACAGAAAGATTCAGAGGAAG 832
578  ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnA 594
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
833  AACACGGGAACCGGATGCCCAACGGGGTAGCTGAGCTCGGTTTGAATCA 882
594  TG 594
      ||
883  GA 884

```

seq_name: gb_gss:AZ356870

seq_documentation_block:

```

LOCUS      AZ356870      644 bp      DNA
DEFINITION 1M0098F09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
ACCESSION  clone UUGC1M0098F09 F, DNA sequence.
VERSION    AZ356870
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 644)
            Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
COMMENT    Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert length: 10000 Std Error: 0.00
            Plate: 0098 Row: F Column: 09
            Seq primer: CGTTGTAAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 644.
FEATURES   location/Qualifiers
            source          1..644
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="UUGC1M0098F09"
                        /clone_1ib="Mouse 10kb plasmid UUGC1M library"
                        /sex="Male"
                        /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                        /note="Vector: PWD42nv; Purified genomic DNA from M.
                        musculus C57BL/6J (male) was obtained from the Jackson
                        Laboratory Mouse DNA Resource
                        (http://www.jax.org/resources/documents/dnares/). The DNA
                        was hydrodynamically sheared by repeated passage through a
                        0.005 inch orifice at constant velocity. The sheared DNA
                        was blunt end-repaired with T4 DNA polymerase and T4
                        polynucleotide kinase. Adaptor oligonucleotides were
                        ligated to the blunt ends in high molar excess. The
                        adaptor DNA was purified and size-selected for a 9.5 to
                        10.5 kb range using preparative agarose gel
                        electrophoresis. Vector DNA was prepared from a derivative
                        of PWD42 (q14732114|gb|AF129072.1), a copy number
                        inducible derivative of plasmid R1. The vector was ligated
                        with adaptors complementary to the insert adaptors and
                        purified. The sheared, adaptor mouse DNA was annealed to
                        adaptor vector DNA, and transformed into
                        chemically-competent E. coli XL10-Gold (Stratagene) cells
                        and selected for ampicillin resistance."
BASE COUNT 185 a 136 c 133 g 190 t
ORIGIN

```

```

alignment_scores:
  Quality: 734.50      Length: 215
  Ratio: 3.866         Gaps: 2
  Percent Similarity: 88.372      Percent Identity: 66.512

```

alignment_block:

US-09-171-553B-10 x AZ356870/rev ..

Align seq 1/1 to reverse of: AZ356870 from: 1 to: 644

```

391 GIUlnAlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTr 407
      ::::::::::::::::::::
639 CAGCTAGTAAAGTGTGATATCTAGTCCGCCCTCAGACACAGATATG 590
      ::::::::::::::::::::
407 pAlacysAsnThrGlyLeuThrProCysValSerThrSerValPheAsn 424
      ::::::::::::::::::::
589 GCGCTTGCAATACCGGCTCCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 540
      ::::::::::::::::::::
424 InsLysAspLeuGluValMetValGlnIleValProArgValTyrTrp 440
      ::::::::::::::::::::
539 GTTCCAAAATTCTGCATTTGTTGCTGCTGCTGCTGCTGCTGCTGCT 490
      ::::::::::::::::::::
441 HisProGluGluValValLeuAspGluTyrAspTyrArgTyrAsnArg 457
      ::::::::::::::::::::
489 CATGATGATAGCTCCTTTTACCAAAATTTGAGATCGGCTC...CGCTG 443
      ::::::::::::::::::::
457 OlyAspArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuG 474
      ::::::::::::::::::::
442 GAGAAGAGAACCCGTTACCTTACCTTGGCAGTTTA..TTAGGATTTGGAG 394
      ::::::::::::::::::::
474 hAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGln 490
      ::::::::::::::::::::
393 TAGCGGCTGAGTACGATACAGAACCGCTGCTTATTATTAAGACCCCA 344
      ::::::::::::::::::::
491 GlnLeuGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAsp 507
      ::::::::::::::::::::
343 TACTATGAAAAA.....CTAGCTGAGCTATGATGTTGATCT 306
      ::::::::::::::::::::
507 uArgAlaLeuLysGluSerValSerAsnLeuGlnGluSerLeuThrSer 524
      ::::::::::::::::::::
305 TAGAACTATAGAACAGCTATATACCAAAATTTAGAAATCTTTAACTTCC 256
      ::::::::::::::::::::
524 euSerGlnValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeu 540
      ::::::::::::::::::::
255 TGTCCGAGTGTGCTACAGATAGAGGGGATTAGATTATTTATCTCT 206
      ::::::::::::::::::::
541 ArgGluGlyGlyLeuGlyAlaAlaLeuLysGluGlyCysPheTyrVal 557
      ::::::::::::::::::::
205 AAAGAGAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156
      ::::::::::::::::::::
557 LAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysL 574
      ::::::::::::::::::::
155 TGACCAATTCAGAGTAAATCAAGATTTCTATGGCCAACTTACAGACGCC 106
      ::::::::::::::::::::
574 euGluArgArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluG 590
      ::::::::::::::::::::
105 TAGATATATAGTAAAGAGAGAGAGAACCAACAGATGTTGTTGAAGC 56
      ::::::::::::::::::::
591 TrpPheAsnArgSerProTyrMetThrThrLeuLeuSerAlaLeu 605
      ::::::::::::::::::::
55 TGGTTTAAATAAAGTCCCTTGGCTGCTGCTGCTGCTGCTGCTGCT 11
      ::::::::::::::::::::
seq_name: gb_gss:A2376517
seq_documentation_block:
LOCUS A2376517 610 bp DNA GSS 02-OCT-2000
DEFINITION IM0130J20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0130J20 F. DNA sequence.
ACCESSION A2376517
VERSION A2376517.1 GI:10490217
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 610)
REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,F., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)

```

```

COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0130 row: J column: 20
Seq primer: CCGTCTAAAGACAGCCGACGAT
Class: plasmid ends
High quality sequence stop: 610.
Location/Qualifiers
1..610
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0130J20"
/clone_lib="Mouse 10kb plasmid library"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1473211419b1AF129072.1) a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 177 a 131 c 120 g 102 t
ORIGIN
alignment_scores:
Quality: 730.50 Length: 204
Ratio: 3.970 Gaps: 2
Percent Similarity: 90.196 Percent Identity: 67.157
alignment_block:
US-09-171-553B-10 x A2376517/rev ..
Align seg 1/1 to reverse of: A2376517 from: 1 to: 610
432 ValGlnIleValProArgValTyrTrpHisProGluGluValValLeuAs 448
      ::::::::::::::::::::
511 CTTTGGCAGTTCTATAGATAGATAGAGTACGCTGAGTATAGATAGAGA 462
      ::::::::::::::::::::
482 ThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuGly 498
      ::::::::::::::::::::
461 ACCGCTGCTTATTAAGACCCCAATATCTATGAA.....GA 424
      ::::::::::::::::::::
498 uLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGluSerValS 515
      ::::::::::::::::::::

```

423	ACTACGCGACGCGATGAGATGTTGACTTATAGAACTATAGAACAGCTATAA	374
515	erAanLeuGIuGIuSerLeuThrSerLeuSerGIuValValLeuGlnAsn	531
373	CCAAATTAGAGAAATCTTTTAACCTCCCTGCGAAGTGGTCTVACAAAT	344
532	ArgArgGIuLeuAspLeuLeuPheLeuArgGIuGIuGIuLeuCysAlaAl	548
323	AGAAAGGGAGTATACCTATTATTTCCTTAAAGAGAGSAGCACTGTGCTGC	274
548	aleuLysGIuGIuCysCysPheTyrValAspHisSerGIuValaIleArg	565
273	CCCTAAAGAGAAATGTTGTTTATGTTAGTGCACCATTCAGAGATATCAAG	224
565	sPSeRmeLaNylsLeuArgLysLysLeuGIuArgArgArgArgGIuArg	581
223	ATTCTATGCGCAACCTTAGAGAACGCGCTGATATACGTAAAGAGAAAGA	174
582	GluAlaAspGIuGIuTyrPheGluGIuGIuTyrPheAsnArgSerProTyrPme	598
173	GAAGGCCAACAAAGAGATGGTTCAAAGCTGGTTTATATAGTCCCTTGCGT	124
598	tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuL	615
123	CACCACTCTCTCTCCACCATAGCAGGAGCCTTGTGATTACACTTATGCTTT	74
615	euleuThrValGIuGIuProCysLeuLysAsnArgPheValAlaIlePheValArg	631
73	TGCTTACTTGTGGCCCTGCATCCTTATTAAGTAGTAGTGTATTATAGA	24
632	GluArgValSer	635
23	GAAGAGATTAAT	12

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0081 row: N column: 06

Seq primer: CACACAGAAACAGCTATGAC

Class: plasmid ends

High quality sequence stop: 632.

FEATURES

SOURCE

1..632

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="Mouse 10kb plasmid UUCGIM library"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD22nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD22 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 203 a 115 c 138 g 176 t
ORIGIN

alignment_scores:

Quality: 713.50 Length: 199
Ratio: 3.920 Gaps: 2
Percent Similarity: 91.457 Percent Identity: 68.844

alignment_block:

US-09-171-553B-10 x AZ814037 ..

Align seg 1/1 to: AZ814037 from: 1 to: 632

```

459 ArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeuGlyThrAl 475
|||||
3 AGAGAACCCGTTACTTACTTGGCAGTTCTATTAGGATTGGAGTAGT 52
475 aValGlyValGlyThrGlyThrAlaValMetLeuGlyProGlnGlnL 492
|||||
53 GGCTGAGTAGTAGTACAGGAGCTGCTTATTAAGACCCCAATACT 102
492 eUGlUuVgSLyLeuGlyLeuHisAlaAlaMetThrGluAspLeuArg 508
|||||
103 ATGAA.....GAACTACGTGACGTATGATGTTGATCTAGA 140
509 AlaleuLysGluSerValSerAsnLeuGluGluSerLeuThrSerLeuSe 525
|||||
141 ACTATAGACAGCTCTATACCAAAATAGAAATCTTAACTTCCTGTC 190
525 rGluValValLeuGluLysAsnArgArgGlyLeuAspLeuLeuPheLeuArg 542
|||||
191 CGAAGTGTGCTACAGATAGAAAGGATTAAGCTTATATCTTAAAG 240
542 LuGlyGlyLeuGlyCysAlaLeuLysGluGlyCysPheTyrValAsp 558

```

```

241 AAGGAGACTCTGTGCTGCCCTAAAGAGATGTTTATTATGTTGAC 290
559 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLysLysLeuG 575
|||||
291 CATTACAGAGTAATCAAGATTTCTATGCGCAAACTTAGAAGCGCTAGA 340
575 uArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrP 592
|||||
341 TATACGTAAAGAGAAAGAGAAAGCCACAGAGATGTTGAAAGCTGCT 390
592 heAsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyPro 608
|||||
391 TTATATAGTCCCTTGGCTCACCACCTCTCTCCACATAGCAGAGACCT 440
609 LeuValValLeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArg 625
|||||
441 TTATATACCTTATGCTTTGCTTACTTTGCTCCTGCAATCCTTATTA 490
625 gPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVal 642
|||||
491 GTTAGTGTGCTTTATTAAGAAAGATTAATGACGTATAGTATGTTAC 540
642 euArgGlnGlnTyrGlnGlyLeuLeuSerGlnGlyGluThrAspLeu 657
|||||
541 TAAAGCAACATATTCG...GTCCCTCAAGAGGTTGAAGAACTGCTC 584

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:18 ; Search time 6550.43 Seconds
(without alignments)
32.809 Million cell updates/sec

Title: US-09-171-553b-13

Perfect score: 20

Sequence: 1 gatggctctctgcccttgg 20

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18.4	92.0	391 11	BF465634	UI-M-CGDP
2	18.4	92.0	429 11	AA797871	VM31h06.r
3	18.4	92.0	435 11	BF565386	UI-R-B01
4	18.4	92.0	444 11	AA472060	V97d05.r
5	18.4	92.0	488 11	BF719561	mab41e06.r
6	18.4	92.0	510 13	AA033199	mi36d04.r
7	18.4	92.0	539 13	AO512916	HS_5139_A
8	18.4	92.0	566 10	AA121599	z078a06.r
9	18.4	92.0	571 10	BE097309	UI-R-B01
10	18.4	92.0	592 10	AM556515	L0269D06-
11	18.4	92.0	600 10	AW488316	UI-M-BH3
12	18.4	92.0	659 10	BE374286	601227645

C 13	18.4	92.0	690 11	BG075579	H3149A02-
14	18.4	92.0	896 11	BG342613	602374796
15	18.4	92.0	2171 12	AK019522	Mus muscu
16	17.4	87.0	257 10	AM326534	19216 MAR
17	17.4	87.0	397 10	AM488995	UI-M-BH3
18	17.4	87.0	455 10	AI013028	EST207479
19	17.4	87.0	581 13	A2777840	2M0012D10
20	17.4	87.0	607 11	W20084	z040f01.r1
21	17.4	87.0	783 13	BH051324	RPCT-24-3
22	17	85.0	691 10	A1735229	at08a01.x
23	17	85.0	1071 10	BE612516	
24	17	85.0	1787 11	BG256691	601452020
25	16.8	84.0	190 11	T25876	BG256691 602370880
26	16.8	84.0	219 10	BB362510	T25876 ESTDIR3 CD3
27	16.8	84.0	250 10	AA478651	BB362510
28	16.8	84.0	268 10	AV228430	AA478651 z019c05.r
29	16.8	84.0	422 13	AO902607	AV228430
30	16.8	84.0	454 11	BF549019	AO902607 LMAJFV1_1
31	16.8	84.0	473 10	AA170346	BF549019 UI-R-A0-a
32	16.8	84.0	492 11	BF554951	AA170346 ms88909.r
33	16.8	84.0	504 13	A2582584	BF554951 UI-R-B0-c
34	16.8	84.0	536 13	AO456957	A2582584 1M0376003
35	16.8	84.0	539 10	BE586605	AO456957 HS_5151_A
36	16.8	84.0	555 10	BE650111	BE586605 WHEP509_G
37	16.8	84.0	558 13	AO523881	BE650111 UI-M-BH3
38	16.8	84.0	576 13	A2730616	AO523881 HS_5246_A
39	16.8	84.0	594 13	A2951110	A2730616 RPCT-24-1
40	16.8	84.0	602 10	A1102103	A2951110 2M0215M20
41	16.8	84.0	610 13	A2432720	A1102103 EST211392
42	16.8	84.0	610 13	A2859488	A2432720 1M0218E17
43	16.8	84.0	654 13	A2741186	A2859488 2M0165D07
44	16.8	84.0	649 13	A2453887	A2741186 RPCT-24-6
45	16.8	84.0	658 13	A2457973	A2453887 1M0255N03
					A2457973 1M0261A07

ALIGNMENTS

RESULT 1
BF465634
LOCUS
DEFINITION
UI-M-CGDP-bq1-h-12-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
UI-M-CGDP-bq1-h-12-0-UI 3', mRNA sequence.
BF465634
ACCESSION
BF465634.1 GI:11534817
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Muskusculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 391)
Normaliztion and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

JOURNAL MEDLINE COMMENT

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mst@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-NO.
Location/Qualifiers

FEATURES

```

source
1. .391
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="UI-M-CG0p-bq1-h-12-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"
BASE COUNT      88 a      100 c      123 g      80 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 11; Length 391;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgctctcctgaccttg 20
|||||
Db 139 GATGCTCTCCTCCCTTTG 158

RESULT 2
AA797871 429 bp mRNA EST 10-FEB-1998
LOCUS      AA797871
DEFINITION w31106.r1 Soares_mammary_gland_NBMNG Mus musculus cDNA clone
ACCESSION  AA797871
VERSION     AA797871.1 GI:2860826
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 429)
AUTHORS    Mairra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
            Geisel, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
            Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
            Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
            Waterston, R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            Contact: Mairra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:659155
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 358.
FEATURES
Source
1. .429
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGC:1245467"
/clone_lib="Soares_mammary_gland_NBMNG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/Note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia)
) with a modified polylinker. Site 1: Not I; Site 2: Eco

```

```

RI; 1st strand cDNA was primed with a Not I - Oligo(dT)
primer [5',
TGTTCACATCTGAGTGGAGCGCGCGGAGGCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      95 a      109 c      128 g      97 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 10; Length 429;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgctctcctgaccttg 20
|||||
Db 13 GATGCTCTCCTCCCTTTG 32

RESULT 3
BF565386 435 bp mRNA EST 12-DEC-2000
LOCUS      BF565386
DEFINITION UI-R-B01-ajv-e-06-0-UI.r1 UI-R-B01 Rattus norvegicus cDNA clone
ACCESSION  BF565386
VERSION     BF565386.1 GI:11675116
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
REFERENCE   1 (bases 1 to 435)
AUTHORS    Bonaldo, M.F., Lennon, G., and Soares, M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            97044477
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LNL (info@image.llnl.gov). IMAGE ID= 1798920
            Seq primer: M13 Forward.
FEATURES
Source
1. .435
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-ajv-e-06-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratseq.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)"

```

BASE COUNT 105 a 121 c 115 g 94 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 435;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 346 GATGCTCTCTCCCTTG 365

RESULT 4

AA472060 444 bp mRNA EST 18-JUN-1997
LOCUS v979d05.r1 Soares_mammary_gland_NbMKG Mus musculus cDNA clone
DEFINITION IMAGE:873897 5' similar to TR:G190352 G190352 PEREGRIN.; mRNA
sequence.

ACCESSION AA472060
VERSION AA472060
KEYWORDS GI:2200051
SOURCE EST.
ORGANISM house mouse.
Mus musculus

REFERENCE 1 (bases 1 to 444)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE The Mashu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513377

Seq primer: -28m13 rev2 EF from Amersham.

FEATURES

Location/Qualifiers

1..444

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:873897"

/clone_lib="Soares_mammary_gland_NbMKG"

/sex="male"

/tissue_type="mammary gland"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia)

RI: 1st strand cDNA was primed with a Not I - Site2: Eco

primer (5'

TGTACCAATCTGAGTGGAGCGCGCGAATGTTT

T 3'): double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia); digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

BASE COUNT 104 a 116 c 127 g 97 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 444;
Best Local Similarity 95.0%; Pred. No. 4.4e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 301 GATGCTCTCTCCCTTG 320

RESULT 5

BF719561 488 bp mRNA EST 03-JAN-2001
LOCUS mab4le06.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
DEFINITION IMAGE:3972947 5' similar to SW:BR14_HUMAN P55201 PEREGRIN.; mRNA
sequence.

ACCESSION BF719561
VERSION BF719561
KEYWORDS GI:12020563
SOURCE EST.
ORGANISM house mouse.
Mus musculus

REFERENCE 1 (bases 1 to 488)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
image.llnl.gov/image/html/resources.shtml
MGI:1472979

Seq primer: -40RP from Gibco
High quality sequence stop: 460.
Location/Qualifiers

FEATURES

1..488

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:3972947"

/clone_lib="Soares_NMEBA_branchial_arch"

/tissue_type="branchial arches"

/dev_stage="embryo, 10.5 dpc"

/lab_host="DH10B (phage resistant)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer (5'

TGTACCAATCTGAGTGGAGCGCGCGAATGTTT

3'): double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia); digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT73 vector. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

BASE COUNT 115 a 135 c 128 g 110 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 488;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gatgctctcgccttg 20
|||||
Db 425 GATGCTCTCTCCCTTG 444

RESULT 6

AA033199 510 bp mRNA EST 22-AUG-1996
LOCUS m136d04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
DEFINITION

TITLE JOURNAL MEDLINE COMMENT

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 439.

FEATURES

source

Location/Qualifiers
1. 566
/organism="Homo sapiens"
/db_xref="GDB:4596797"
/db_xref="taxon:9606"
/clone="IMAGE:564274"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 114 a 157 c 141 g 151 t 3 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 566;
Best Local Similarity 95.0%; Pred. No. 4.5e-02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gatggctctcctgccttgg 20
|||||
Db 348 GAGGGCTCTCCTGCCCTTGG 329

RESULT 9

BE097309/c
LOCUS BE097309 571 bp mRNA EST 12-JUN-2000
DEFINITION UI-R-B01-apv-g-12-0-UI.sl UI-R-B01 Rattus norvegicus cDNA clone
UI-R-B01-apv-g-12-0-UI 3', mRNA sequence.
ACCESSION BE097309
VERSION BE097309.1 GI:8488240
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 571)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the

normalized medulla library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1. 571
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-apv-g-12-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-B01
TAG_TISSUE=medulla
TAG_SEQ=GAACCG"

BASE COUNT 143 a 152 c 134 g 141 t 1 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 571;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gatggctctcctgccttgg 20
|||||
Db 528 GATGGCTCTCCTGCCCTTGG 509

RESULT 10

AW556515/c
LOCUS AW556515 592 bp mRNA EST 31-AUG-2000
DEFINITION L0269D06-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
clone L0269D06 3', mRNA sequence.
ACCESSION AW556515
VERSION AW556515.1 GI:7201944
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 592)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE 20381348
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: L0269 row: D column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 592
POLYA=Yes.
Location/Qualifiers
1. 592
/organism="Mus musculus"

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0269D06"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/notes="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: Sali; Site_2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from GibcoBRL]
15'-pGACTAGTCTAGATCCGAGCGCCGCTTTTCTTTT-3'] from
2.5ug of total RNA . The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-sal3 (include Sali sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with Sali and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into Sali/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao." 3 others
BASE COUNT      140 a 166 c 146 g 137 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 592;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 gatggctctctgcctttg 20
|||||
Db  511 GATGGCTCTCTCTCCCTTTG 492

```

```

RESULT 11
AW488316/C
LOCUS      AW488316      600 bp      mRNA      EST      24-FEB-2000
DEFINITION UI-M-BH3-arp-b-06-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION  AW488316
VERSION    AW488316.1 GI:7058586
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 600)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            9704477
            Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: m3ST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized cerebellum library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: Researchers may obtain BMAP cDNA clones

```

from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA-Yes.

```

FEATURES             Location/Qualifiers
     source            1..600
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-arp-b-06-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /notes="Vector: pT73D-Pac (Pharmacia) with a modified
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=cerebellum
TAG_SEQ=GACTC"
BASE COUNT      141 a 166 c 143 g 150 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 600;
Best Local Similarity 95.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 gatggctctctgcctttg 20
|||||
Db  528 GATGGCTCTCTCTCCCTTTG 509

RESULT 12
BE374286
LOCUS      BE374286      659 bp      mRNA      EST      21-JUL-2000
DEFINITION 601272645F1 NCL_CGAP_Maml Mus musculus cDNA clone IMAGE:3585899 5',
            mRNA sequence.
ACCESSION  BE374286
VERSION    BE374286.1 GI:9319649
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 659)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Found through the MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM8746 row: j column: 12
 High quality sequence stop: 586.
 Location/Qualifiers

FEATURES

source

1. .659
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3585899"
 /clone_lib="NCI_CGAP_Maml"
 /dev_stage="3 months, virgin"
 /tissue_type="tumor, biopsy sample"
 /lab_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT
 ORIGIN

149 a 178 c 189 g 143 t

Query Match

Best Local Similarity 92.0%; Score 18.4; DB 10; Length 659;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatggctctcctgccttg 20

||||| ||||| ||||| |||||

Db 452 GATGGCTCTCTCCCTTTG 471

RESULT 13

BG075579/c

LOCUS

DEFINITION H3149A02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone

H3149A02 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

690 bp mRNA EST 26-JAN-2001
 BG075579 H3149A02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3149A02 3', mRNA sequence.
 BG075579
 BG075579.1 GI:12558148
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 690)
 T.S., Carter, M.G. and Ko, M.S.H.
 Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
 Verification and initial annotation of NIA mouse 15K cDNA clone set
 Unpublished (2001)
 Other_ESTs: H3149A02-5
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdaei@gsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
 Plate: H3149 row: A column: 02
 Seq primer: -21M13 Forward
 High quality sequence stop: 690
 POLYA=Yes.
 Location/Qualifiers

FEATURES

source

1. .690
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="H3149A02"

/clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
 clone is among a rearranged set of 15,247 clones from 11
 embryo cDNA libraries (including preimplantation stage
 embryos from unfertilized egg to blastocyst, embryonic
 part of E7.5 embryos, extraembryonic part of E7.5 embryos
 , and E12.5 female mesonephros/gonad) and one newborn
 ovary cDNA library. Average insert size 1.5 kb. All
 source libraries are cloned unidirectionally with Oligo(dT
)-Not primers. References include: (1) Genome-wide
 expression profiling of mid-gestation placenta and embryo
 using a 15,000 mouse developmental cDNA microarray, 2000,
 Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
 Large-scale cDNA analysis reveals phased gene expression
 patterns during preimplantation mouse development, 2000,
 Development, 127: 1737-1749; (3) Genome-wide mapping of
 unselected transcripts from extraembryonic tissue of
 7.5-day mouse embryos reveals enrichment in the t-complex
 and under-representation on the X chromosome, 1998, Hum
 Mol Genet 7: 1967-1978."
 Mol Genet 7: 1967-1978."
 BASE COUNT 158 a 197 c 170 g 165 t
 ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 690;
 Best Local Similarity 95.0%; Pred. No. 4.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatggctctcctgccttg 20

||||| ||||| ||||| |||||

Db 512 GATGGCTCTCTCCCTTTG 493

RESULT 14

BG342613

LOCUS

DEFINITION BG342613 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4482310 5',
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

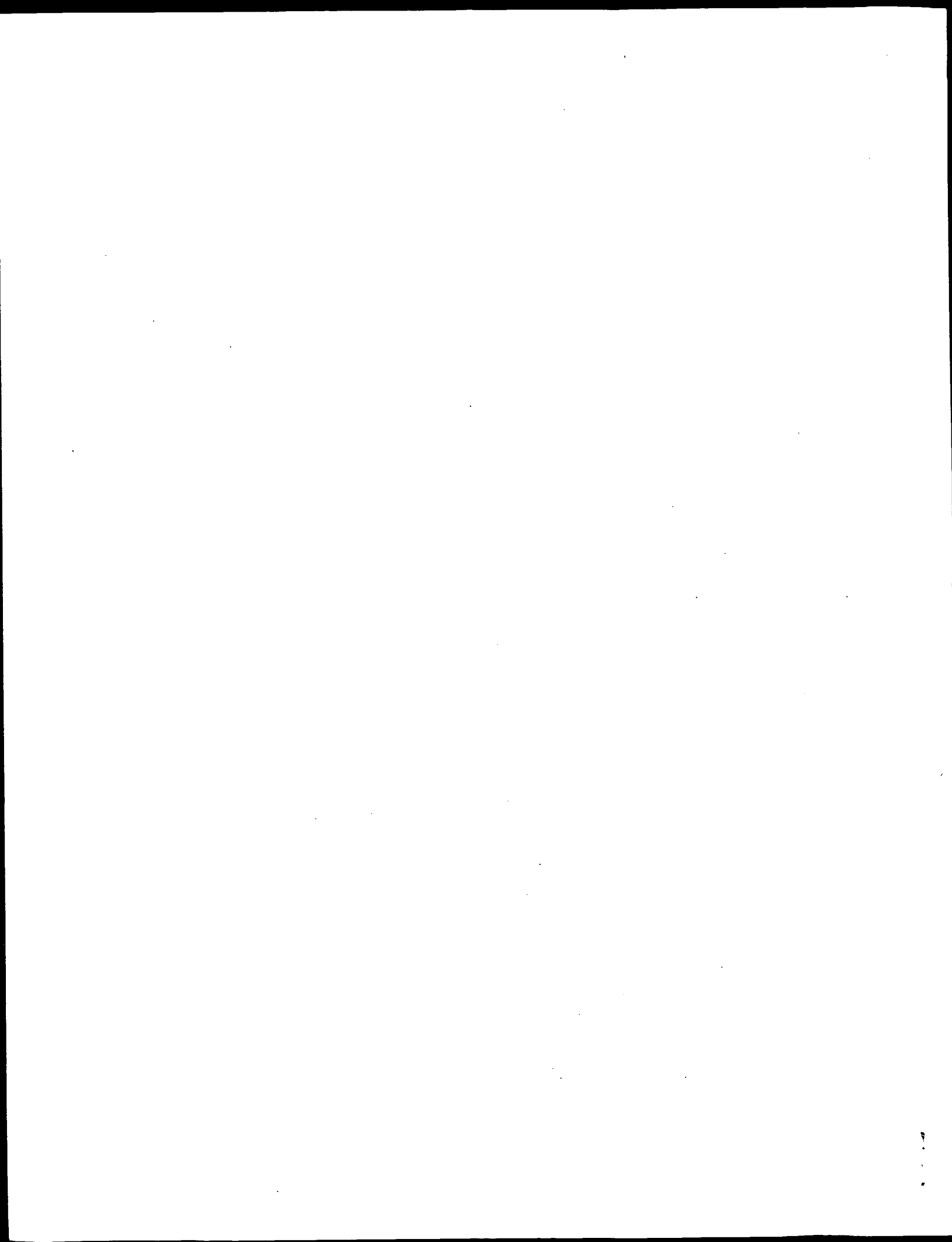
JOURNAL

COMMENT

896 bp mRNA EST 27-FEB-2001
 BG342613 602374796f1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4482310 5',
 mRNA sequence.
 BG342613
 BG342613.1 GI:13149051
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 896)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10318 row: p column: 23
 High quality sequence stop: 564.
 Location/Qualifiers

FEATURES
 source

1. .896
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4482310"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: PCMV-SPORT6; Site_1: NotI;



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2002, 23:37:21 : Search time 6550.43 Seconds
(without alignments)
29.528 Million cell updates/sec

Title: US-09-171-553B-14

Perfect score: 18

Sequence: 1 ccacagtcgtacaccacg 18

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estov:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	18	100.0	558	11	BI360879 38957 MA
C 2	17	94.4	323	13	AZ056887 RPCI-23-4
C 3	17	94.4	559	13	AZ749422 RPCI-24-1
C 4	17	94.4	655	13	BH097469 RPCI-24-3
C 5	16.4	91.1	155	10	AW346676 29231 MAR
C 6	16.4	91.1	166	10	AW346655 29171 MAR
C 7	16.4	91.1	232	10	AV312947 AV312947
C 8	16.4	91.1	254	10	AV289929 AV289929
C 9	16.4	91.1	261	11	BI184980 UNL-P-PN-
C 10	16.4	91.1	290	13	AZ113096 RPCI-23-4
C 11	16.4	91.1	294	11	BF713588 MI-P-O2-a
C 12	16.4	91.1	300	11	BI184279 UNL-P-PN-

C 13	16.4	91.1	370	13	AZ313323
C 14	16.4	91.1	379	13	AZ793176
C 15	16.4	91.1	395	13	AZ831760
C 16	16.4	91.1	425	13	AZ036915
C 17	16.4	91.1	427	13	AZ58978 RPCI-23-1
C 18	16.4	91.1	432	13	AZ909325 RPCI-24-1
C 19	16.4	91.1	458	11	BI185480 UNL-P-PN-
C 20	16.4	91.1	468	11	BI182727 UNL-P-PN-
C 21	16.4	91.1	468	11	BI186066 UNL-P-PN-
C 22	16.4	91.1	470	13	AZ387701
C 23	16.4	91.1	495	13	AZ955800
C 24	16.4	91.1	499	11	BF704354 MI-P-O3-a
C 25	16.4	91.1	513	10	AW786160
C 26	16.4	91.1	521	11	BI399852 MI-P-AV1-
C 27	16.4	91.1	552	13	AZ242483 RPCI-23-7
C 28	16.4	91.1	553	13	AZ999118
C 29	16.4	91.1	579	11	BI183723 UNL-P-PN-
C 30	16.4	91.1	602	11	BF712162 MI-P-E6-a
C 31	16.4	91.1	607	13	BH088353 RPCI-24-2
C 32	16.4	91.1	613	13	AZ454083
C 33	16.4	91.1	615	13	AZ385535
C 34	16.4	91.1	622	13	AZ115761 RPCI-23-4
C 35	16.4	91.1	655	13	AZ836447
C 36	16.4	91.1	681	13	AZ115038 RPCI-23-4
C 37	16.4	91.1	683	11	BG122508
C 38	16.4	91.1	767	13	BH125701
C 39	16.4	91.1	864	11	BF790922
C 40	16.4	91.1	1042	10	BE544440
C 41	16.4	91.1	2260	12	AK015210 Mus muscu
C 42	16	88.9	686	11	BF501296
C 43	15.4	85.6	142	13	AZ601453
C 44	15.4	85.6	252	11	W66405
C 45	15.4	85.6	259	13	AZ308532

ALIGNMENTS

RESULT 1

BI360879/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI360879 558 bp mRNA EST 01-AUG-2001
38957 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.
BI360879
BI360879.1 GI:15056907
EST.
Pig.
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 558)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980504 e. Vector identified by cross_match with the -mismore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 142 row: B column: 4
Seq primer: ATTAGGTGACATATAG.
Location/Qualifiers
1..558
/organism="Sus scrofa"


```

/db_xref="taxon:9823"
/clone_lib="MARC 2Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      129 a 123 c 119 g 187 t
ORIGIN

Query Match      100.08; Score 18; DB 11; Length 558;
Best Local Similarity 100.08; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 18
|||||
Db 178 CCACAGTCGTACACCACG 161

RESULT 2
LOCUS      AZ056887      323 bp      DNA      GSS      30-MAR-2000
DEFINITION RPCI-23-428K7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-428K7,
DNA sequence.
ACCESSION  AZ056887
VERSION    AZ056887.1 GI:7348123
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 323)
AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE     Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 428 row: K column: 7
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
source      1..323
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-428K7"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      83 a 59 c 72 g 109 t
ORIGIN

```

```

Query Match      94.48; Score 17; DB 13; Length 323;
Best Local Similarity 100.08; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
Db 218 CCACAGTCGTACACCAC 234

RESULT 3
LOCUS      AZ749422      559 bp      DNA      GSS      25-JAN-2001
DEFINITION RPCI-24-115E19.TV RPCI-24 Mus musculus genomic clone RPCI-24-115E19
, DNA sequence.
ACCESSION  AZ749422
VERSION    AZ749422.1 GI:12534581
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 559)
AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Rusell,D., de Jong,P. and Fraser,C.M.
TITLE     Mouse BAC End Sequences from Library RPCI-24
JOURNAL   Unpublished (1999)
COMMENT   Other GSSs: RPCI-24-115E19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 115 row: E column: 19
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
source      1..559
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-115E19"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      133 a 125 c 137 g 164 t
ORIGIN

Query Match      94.48; Score 17; DB 13; Length 559;
Best Local Similarity 100.08; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccac 17
|||||
Db 411 CCACAGTCGTACACCAC 395

RESULT 4
LOCUS      BH097469

```

LOCUS
DEFINITION RPCI-24-315L22.TV RPCI-24 Mus musculus genomic clone RPCI-24-315L22
ACCESSION BH097469
VERSION BH097469
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 655)
Tsdaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P., and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-24-315L22.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 315 row: L column: 22
Seq primer: T7
Class: BAC ends.

FEATURES

source
1..655
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-315L22"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; site_1: BamHI; site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 180 a 157 c 148 g 170 t
ORIGIN

Query Match 94.4%; Score 17; DB 13; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

QY 1 ccacagtcgtacaccac 17
|||||
Db 116 CCACAGTCGTACACCAC 132

RESULT 5
AW346676/c
LOCUS AW346676 155 bp mRNA EST 09-JUL-2000
DEFINITION 29231 MARC LP1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW346676
VERSION AW346676.1 GI:6844386
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 155)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.

TITLE and Keele,J.W.
JOURNAL Design and use of two pooled tissue normalized cDNA libraries for
COMMENT EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 13 row: B column: 12
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..155
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC LP1G"
/tissue_type="pooled"
/lab_host="DHI0B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 34 a 45 c 39 g 37 t
ORIGIN

FEATURES

source
1..155
Location/Qualifiers
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC LP1G"
/tissue_type="pooled"
/lab_host="DHI0B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 34 a 45 c 39 g 37 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 10; Length 155;
Best Local Similarity 94.4%; Pred. No. 3.5e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1;

QY 1 ccacagtcgtacaccac 18
|||||
Db 112 CCACAGTCGTACGCCACG 95

RESULT 6
AW346655/c
LOCUS AW346655 166 bp mRNA EST 09-JUL-2000
DEFINITION 29171 MARC LP1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW346655
VERSION AW346655.1 GI:6844365
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 166)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for
JOURNAL EST discovery in swine
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 13 row: H column: 6
Seq primer: ATTTAGGTGACACTATAG.

FEATURES

source

```

Location/Qualifiers
1. .166
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled thymus from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      45 a      45 c      39 g      37 t
ORIGIN
1. .232
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5830407B15"
/sex="male"
/tissue_type="thymus"
/dev_host="adult"
/Note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGATCCGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATAATTAATCCCTCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI."
BASE COUNT      49 a      64 c      39 g      80 t
ORIGIN

```

```

Query Match      91.1%; Score 16.4; DB 10; Length 166;
Best Local Similarity 94.4%; Pred. No. 3.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ccacagtcgtacaccacg 18
||||| ||||| |||||
Db 112 CCACAGTCGTAGCCACG 95

```

RESULT 7

```

AV312947/c      232 bp      mRNA      EST      08-NOV-1999
LOCUS      AV312947 RIKEN full-length enriched, adult male thymus Mus musculus
DEFINITION      cDNA clone 5830407B15 3', mRNA sequence.
ACCESSION      AV312947
VERSION      AV312947.1 GI:6278199
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus.

```

```

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 232)

```

```

AUTHORS
Konno,H., Aizawa,K., Akahira,S., Hayatsu,N., Hirozane,T., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Izawa,M., Kadota,K., Kagawa,I., Kai
Ishii,Y., Ishikawa,T., Itoh,M., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
Y.

```

TITLE

JOURNAL

COMMENT

```

Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

```

FEATURES

Location/Qualifiers

TITLE

JOURNAL

COMMENT

```

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

```

ORGANISM	Sus scrofa
REFERENCE	1 (bases 1 to 261)
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS

and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-475F22.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 475 row: F column: 22
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1. .290
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-23-475F22"
 /clone="RPCI-23-475F22"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 84 a 59 c 84 g 63 t
 ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 290;
 Best Local Similarity 94.4%; Pred. No. 4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ccacagtcgtacacacg 18
 ||||| |||||
 Db 191 CCACAGTCATACACCAG 208

RESULT 11
 BF713588
 LOCUS 294 bp mRNA EST 02-JAN-2001
 DEFINITION MI-P-02-adv-e-01-1-UM.s1 MI-P-02 Sus scrofa cDNA clone
 MI-P-02-adv-e-01-1-UM 3', mRNA sequence.
 ACCESSION BF713588
 VERSION BF713588.1 GI:12013063
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 294)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tugle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401

Email: cktugle@iastate.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1. .294
 /organism="Sus scrofa"
 /strain="crossbred"
 /db_xref="taxon:9823"
 /clone="MI-P-02-adv-e-01-1-UM"
 /clone_lib="MI-P-02"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: EcoRI; The MI-P-02 library is derived from ovary at estrus day 5. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.
 TAG_SEQ=None found"

BASE COUNT 69 a 61 c 77 g 87 t
 ORIGIN

Query Match 91.1%; Score 16.4; DB 11; Length 294;
 Best Local Similarity 94.4%; Pred. No. 4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ccacagtcgtacacacg 18
 ||||| |||||
 Db 62 CCACAGTCGTACGCCAG 79

RESULT 12
 BI184279

LOCUS 300 bp mRNA EST 10-JUL-2001
 DEFINITION UNL-P-FN-CK-g-02-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
 UNL-P-FN-CK-g-02-0-UNL 3', mRNA sequence.
 ACCESSION BI184279
 VERSION BI184279.1 GI:14658688
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 300)
 AUTHORS Caetano,A.R., Johnson,R.K. and Pomp,D.
 TITLE Generation and sequence characterization of a normalized cDNA library from swine ovarian follicles
 JOURNAL Unpublished (2001)
 COMMENT Contact: Pomp, D
 Department of Animal Science
 University of Nebraska, Lincoln
 Lincoln, NE 68583-0908, USA
 Tel: 402 472 6416
 Fax: 402 472 6362
 Email: dpomp@unl.edu

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail.
 Seq primer: M13 -29
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1. .300
 /organism="Sus scrofa"
 /strain="University of Nebraska, Lincoln Swine Selection Lines"

/db_xref="taxon:9823"
/clone="UNL-P-FN-ck-g-02-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996.

TAG_SEQ=None found"
69 a 62 c 71 g 97 t 1 others

Query Match 91.1%; Score 16.4; DB 11; Length 300;
Best Local Similarity 94.4%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
||||| |||||||
DB 59 CCACAGTCATACACCACG 76

RESULT 13

AZ313323/c
LOCUS AZ313323 370 bp DNA GSS 29-SEP-2000
DEFINITION IM0029A06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0029A06 R, DNA sequence.

ACCESSION AZ313323
VERSION AZ313323.1 GI:10358137
KEYWORDS GSS
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 370)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0029 row: A column: 06
Seq primer: CCACAGGAACAGCATATGACC
Class: plasmid ends
High quality sequence stop: 370.

FEATURES
source
1. .370
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0029A06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWB42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 66 a 100 c 84 g 120 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 370;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccacg 18
||||| |||||||
DB 165 CCACGGTCGTACACCACG 148

RESULT 14

AZ793176
LOCUS AZ793176 379 bp DNA GSS 16-FEB-2001
DEFINITION 2M0046N11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0046N11 F, DNA sequence.

ACCESSION AZ793176
VERSION AZ793176.1 GI:12937866
KEYWORDS GSS
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 379)
REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0046 row: N column: 11
Seq primer: CGTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 379.

FEATURES
source
1. .379
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0046N11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 120 a 80 c 95 g 84 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 379;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccag 18
||||||| |||||||
Db 75 CCACAGTCATACACCAG 92

RESULT 15

AZ831760 395 bp DNA GSS 20-FEB-2001
LOCUS 2M011K09R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M011K09 R, DNA sequence.

ACCESSION AZ831760
VERSION AZ831760.1 GI:13001668
KEYWORDS GSS.
SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 395)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0111 row: K column: 09
Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends
High quality sequence stop: 395.

FEATURES

source
1..395
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M011K09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 121 a 88 c 112 g 74 t
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 395;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacagtcgtacaccag 18
||||||| |||||||
Db 253 CCACAGTCATACACCAG 270

Search completed: February 23, 2002, 23:37:23
Job time: 9444 sec

